Optimal Energy Management Strategies for Energy Internet via Deep Reinforcement Learning Approach

Haochen Hua[a], Yuchao Qin[a], Chuantong Hao[a], Junwei Cao[a,*]

[a]*Research Institute of Information Technology, Tsinghua University, Beijing, China*

Abstract

This paper investigates the energy management problem in the field of energy Internet (EI) with interdisciplinary techniques. The concept of EI has been proposed for a while. However, there still exist many fundamental and technical issues that have not been fully investigated. In this paper, a new energy regulation issue is considered based on the operational principles of EI. Multiple targets are considered along with some constraints. Then, the practical energy management problem is formulated as a constrained optimal control problem. Due to its complexity, the problem considered in this paper cannot be simply solved by conventional methods. To obtain the desired control scheme, a model free deep reinforcement learning algorithm is applied. A practical solution is obtained, and the feasibility as well as the performance of the proposed method are evaluated with numerical simulations.

Keywords: Energy Internet, Energy Routers, Microgrids, Optimal Control, Deep Reinforcement Learning

1. Introduction

As alternative to conventional fossil fuels, the demand for renewable energy has considerably increased during the past decades. As such, investigation on renewable power generation, e.g, solar power and wind power have attracted much attention [1, 2]. Although renewable energy sources (RESs) have advantages including sustainable and environmental friendly, they have inherent defects such as nonlinear, intermittent and stochastic [3, 4]. On the other hand,

*Corresponding author

microgrids (MGs) have been viewed as a solution to the challenges facing traditional power systems [5, 6]. When vast distributed RESs are utilized in MGs, it is difficult to achieve a reliable power balance in MGs (especially the isolated ones), if without proper regulation; see, e.g., [7–10].

In recent years, to solve the aforementioned challenges, research emphasis has been directed towards the development of energy Internet (EI) which was first proposed in [11]. Inspired by the core of Internet, the EI treats MGs as infrastructures at the end of future energy systems, allowing the access of large amounts of distributed energy resources (DERs) [12, 13]. In [14], it is pointed out that EI can be viewed as the upgraded version of the smart grid. A variety of networking topology of EI has been introduced in [15]. Within the scope of EI, multiple MGs are interconnected via energy routers (ERs) [16, 17], also known as energy hubs [18], or power routers [19]. In this fashion, energy exchange can be realized via the interconnected MGs, and the capacity of their energy storage (ES) devices can be shared, such that power generation-consumption balance for the whole EI scenario can be achieved. According to [12–15], the basic energy management principle in EI is that autonomous power balance in single MG should be achieved with priority. If local MG's power balance is difficult to be achieved, then energy exchange in wide area network shall be implemented.

In the field of EI, research on energy control strategies has attracted much attention and significant advances on this topic have been made; see, e.g., [20]-[23]. In [20], voltage regulation issue for one DC MG in EI scenario has been transformed into a non-fragile robust H_∞ control problem. Besides, in the field of EI, H_∞ control theory has been applied to regulate the frequency deviations in AC MGs [21]. A class of distributed coordinated control algorithm for EI has been proposed in [22]. A graph theory based energy routing algorithm in EI has been studied in [23].

It is notable that most of the control problems in power systems are solved based on explicit mathematical models of various electrical devices. For example, ordinary differential equations (ODEs) are used to represent the power dynamics of photovoltaic (PV) units and wind turbine generators (WTGs) and

loads in e.g., [4, 9, 24, 25], while stochastic differential equations (SDEs) [26] are used to represent the power dynamics of RESs and loads in e.g., [10, 20, 21, 27]. Although the SDEs can reflect the stochastic nature of the DERs, it is difficult for engineers to obtain their accurate mathematical models. It is notable that in order to represent power dynamics for a relatively long time period (for example, one day), a mathematical model with complicated differential equations shall be established, which is somehow restrictive. In this sense, finding a series of mathematical models for the power of DERs in EI is time-consuming as well as costly.

On the other hand, the applications of artificial intelligence on power systems has been popular in the past decade. The electricity forecasting is one of the most important issues for EI. There are already a number of literatures on the electricity forecasting for PVs, WTGs, loads, etc.; see, e.g., [28–30]. To illustrate, neural networks are used for the power modeling of PVs and loads in [28] and [29], respectively. Based on extreme learning machine and improved gravitational search algorithm, a novel short-term load forecasting method has been proposed in [30]. Besides, for the application of reinforcement learning into residential load control, readers can refer to [31]. A novel distributed energy management approach based on deep learning algorithm has been reported in [32]. Since the estimation performance of the advanced methods in these research outputs are satisfactory and most of these techniques are practical, it is feasible to design control schemes for the EI system based on the power forecast results.

In this paper, the energy management problem for a typical scenario of EI is investigated. A generalized EI scenario is considered, in which multiple MGs are interconnected via ERs. Each MG is assumed to consist of PV units, WTGs, micro-turbines (MTs), diesel engine generators (DEGs), battery energy storage (BES) devices and loads. Historical data from [33] are used as the forecast results for power of PVs, WTGs, and loads for simplicity. Based on the energy management principle of EI, the desired targets for optimal energy management are formulated as cost functions mathematically. Next, a series of

penalty functions are formulated. Besides, some constraints for the optimization problem are introduced. Next, the energy management issue considered in this paper is formulated as an optimal control problem.

Generally, the Hamilton-Jaccobi-Bellman (HJB) equation is used to find the solution to the continuous/discrete time optimal control problem [34]. For the discrete time system, it is usually called Bellman equation. There have been many algorithms for the optimal control problem based on Bellman equation; see e.g., [35, 36]. However, these methods cannot be applied to solve the optimal control problem formulated in this paper, the reasons of which are analysed below.

Firstly, most of the existing solutions to the HJB equations adopt "grid based" methods, which means that they rely on the discretization of action space and state space. As a result, these methods suffer from the "curse of dimensionality". The computation and storage complexities increase exponentially with the growth of the dimensions of action space and state space. Although there are a few approaches providing polynomial-time solutions [37], they may rely on some specified property of the problem. In this paper, the considered EI system is rather complex. There is no system modelling for the power of PVs, WTGs, and loads. Their power dynamics are just assumed to be time series data obtained from proper electricity estimation techniques. Hence, there is no explicit formula for these time series. Thus, the conventional methods mentioned above cannot be applied in this paper.

With the development of the reinforcement learning theory and algorithm, the solvability to a general optimal control problem becomes possible. In this paper, we convert our considered optimal control issue into a reinforcement learning problem which can be solved by the A3C algorithm [38] The importance and contribution of this paper can be highlighted as follows.

- Optimal energy management strategies are considered for a *generalized* EI system, allowing for a variety of optimization targets. The considered objectives include the transmission loss for ERs, power generation cost

for MTs and DEGs, and lifetime extension for BES devices. Different kinds of trade-off between these objectives can be achieved by adjusting their weighting factors. It is notable that the above targets have not been considered *simultaneously* in EI scenarios.

- By intelligently scheduling the energy flow of multiple MGs and ERs, the power supply-demand balance is realized not only in each individual MG, but also in the entire EI system, such that the customers can benefit from the guaranteed reliable power supply.

- The power of PVs, WTGs and loads are represented with data directly, based on which, a new energy optimization problem is considered. A model free approach is applied to solve the problem. In this sense, the system modelling error is successfully avoided, thus making the obtained control strategies more reliable.

- When formulating the cost functions, a class of penalty functions are considered for the constraints of the EI system. The rational utilization of MTs, DEGs, ERs and BES devices are considered. In this sense, the energy management approach proposed in this paper is of both theoretical complexity and practical usefulness.

- In this paper, we consider control problems among cross disciplinary subjects, including mathematics, computer sciences and smart grids. Since the formulated problem is complicated, in the sense that it cannot be effectively solved by conventional methods, e.g., particle swarm optimization (PSO) [39], genetic algorithm (GA) [40], simulate anneal arithmetic (SAA) [41], etc., we apply the new deep reinforcement learning approach to solve the synthetical optimal control problem. The most recent A3C algorithm is applied to achieve the target. The simulation results show the effectiveness of the proposed method.

The rest of this paper is organized as follows: Section 2 introduces the EI system modelling. The optimal control problem formulation is introduced in

Section 3. In Section 4, solution to the energy management issue is provided. Numerical examples are illustrated in Section 5. Finally, Section 6 concludes the paper.

2. System modelling

As is shown in Figure 1, the EI network is assumed to be disconnected with the power utility. Each MG in the considered EI system is interconnected via ERs. Each ER in the system is able to exchange electric power with other ERs through the power transmission lines. All of the MGs are assumed to consist of the same components, including PVs, WTGs, MTs, DEGs, BES devices and loads. The structure of such MG is presented in Figure 2.

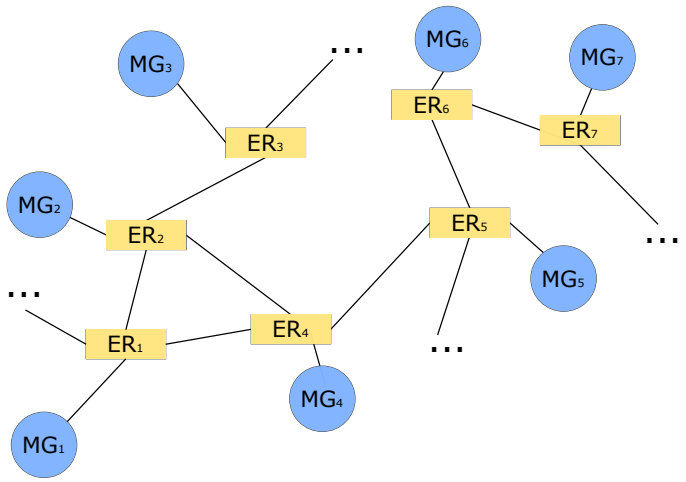

Figure 1: EI topology

In this paper, historical data from [33] are used as the power forecast results of PVs, WTGs and loads. These data are sampled at 1/60 Hz, so the power of PVs, WTGs and loads in the MGs are represented with discrete time series with time step of 1 minute.

Suppose that there are totally N MGs and N ERs in the considered EI system. The subscripts of ERs belong to the set $V = \{1, 2, \ldots, N\}$. We denote

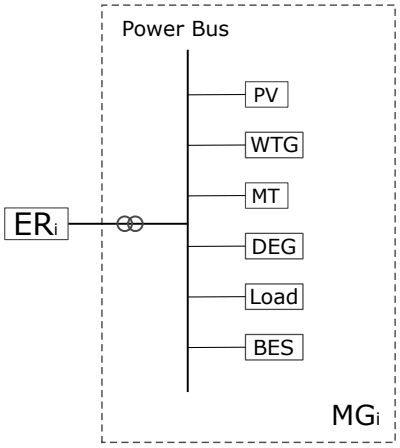

Figure 2: MG topology

the ith MG as $MG_i, i \in V$ and denote the ER tied to MG_i as $ER_i, i \in V$. The set of the connections among ERs is denoted as E. We have

$$E = \{(i,j)|ER_i \leftrightarrow ER_j, \quad i,j \in V\},$$

where $\leftrightarrow$ means that ER_i and ER_j are interconnected. Thus, the total number of the connections is $\frac{1}{2}|E|$. In this section, time t in the power of DERs and ERs is omitted for notation simplicity. For every two ERs, ER_i and ER_j in the system, the energy transmitted from ER_i to ER_j is denoted as $P_{i,j}^{ER}$. With these notations, we have

$$\begin{aligned} P_{i,j}^{ER} &= -P_{j,i}^{ER}, \quad i,j \in V, \\ P_{i,j}^{ER} &= 0, \quad (i,j) \notin E, \\ P_{i,i}^{ER} &= 0, \quad i \in V, \end{aligned}$$

where $P_{i,j}^{ER} \geq 0$ means that the energy is transmitted from ER_i to ER_j, and vice versa.

In real power systems, the capacities of power transmission lines are affected by a variety of factors, e.g., length of the line, temperature [42, 43]. Hence, there exist an upper bound for the power transmitted through a power transmission

line. We denote such upper bound for the transmission line between MG_i and MG_j as $U_{i,j}^{ER}$. Apparently, $U_{i,j}^{ER} = U_{j.i}^{ER}$, and (1) is established.

$$0 \leq \left|P_{i,j}^{ER}\right| \leq U_{i,j}^{ER}, \quad i,j \in V. \tag{1}$$

Here, the $|\cdot|$ stands for the absolute value function.

The power of PVs, WTGs and loads are considered to be uncontrollable, but could be forecasted with a certain degree of accuracy. In MG_i, the power forecast results for PVs, WTGs, and loads are denoted as P_i^{PV}, P_i^{WTG} and P_i^L, respectively. The sum of the power for these uncontrollable components are denoted as P_i^{UC} which is assumed to be obtained by

$$P_i^{UC} = P_i^L - P_i^{PV} - P_i^{WTG} + P_i^E,$$

where P_i^E is a scalar Weiner process [26]. Due to the stochastic and uncertain nature of PVs, WTGs and loads, there is no doubt that P_i^{UC} has similar stochastic characteristics. The Weiner process P_i^E is used to represent such character. We denote the output power of MTs, DEGs, and the power transmitted to ER_i as P_i^{MT}, P_i^{DEG} and P_i^{ER}, respectively. According to the notations for $P_{i,j}^{ER}$, we have

$$P_i^{ER} = \sum_{j \in V} P_{j,i}^{ER},$$

where $P_i^{ER} \geq 0$ corresponds to the situation that MG_i absorbs energy from other MGs; $P_i^{ER} \leq 0$ means that MG_i transmits energy to other MGs.

In each MG, the output power of MTs and DEGs is controlled by the EI system manager. Usually the control decisions are generated according to the system states and the pre-set control schemes. Generally, both MTs and DEGs have their maximum output power. For output power of MTs and DEGs in MG_i, the following constraints are applied,

$$0 \leq P_i^{MT} \leq U_i^{MT}, \quad i \in V,$$
$$0 \leq P_i^{DEG} \leq U_i^{DEG}, \quad i \in V,$$

where U_i^{MT} and U_i^{DEG} are the upper bounds for power of MTs and DEGs, respectively.

The charge/discharge power and state of charge (SOC) for BES devices in MG_i are denoted as P_i^{BES} and SOC_i. The BES devices are used to balance the power generation and consumption in MGs, which means that BES devices can passively absorb the power deviations in MGs. It might happen that instant power deviation in a MG is too large for the BES devices. In order to protect BES devices from being damaged, their maximum charge/discharge power is restricted. Meanwhile, the SOC should also be maintained within a proper range. The running constraints for BES devices are given by

$$\begin{aligned} 0 \leq \left|P_i^{BES}\right| \leq U_i^{BES}, \quad i \in V, \\ L_i^{SOC} \leq SOC_i \leq U_i^{SOC}, \quad i \in V, \end{aligned}$$

where U_i^{BES} is the maximum allowed charge/discharge power for BES devices; L_i^{SOC} and U_i^{SOC} are the lower and upper bounds for SOC, respectively.

Since the maximum charge/discharge power of BES devices is restricted, an inappropriate control policy may lead to the unbalanced supply-demand power in one MG, although BES devices have been fully utilized. To deal with such problem, the slack variable P_i^{UB} is introduced in (2).

$$P_i^{BES} = P_i^{UC} - P_i^{ER} - P_i^{MT} - P_i^{DEG} - P_i^{UB}. \tag{2}$$

The slack variable P_i^{UB} is obtained with the following formula,

$$P_i^{UB} = \begin{cases} 0, & |\Delta P_i| \leq U_i^{BES}, \\ \Delta P_i - U_i^{BES}, & \Delta P_i \geq U_i^{BES}, \\ \Delta P_i + U_i^{BES}, & \Delta P_i \leq -U_i^{BES}, \end{cases}$$

where $\Delta P_i = P_i^{UC} - P_i^{ER} - P_i^{MT} - P_i^{DEG}$. During the operation of the MG system, P_i^{UB} should be kept to be zero, such that the unbalanced power deviations in MGs could be absorbed by BES devices completely. In this sense, an autonomous operation of MG in EI can be achieved.

According to [44], the dynamics of SOC are given in (3),

$$\dot{SOC}_i = -\eta_i P_i^{BES}/Q_i, \tag{3}$$

where Q_i is the capacity of BES devices; η_i is the charge/discharge coefficient for BES devices and it is defined in (4).

$$\eta_i \triangleq \begin{cases} \eta_i^{in}, & P_i^{BES} \leq 0, \\ 1/\eta_i^{out}, & P_i^{BES} \geq 0. \end{cases} \tag{4}$$

The coefficients η_i^{in} and η_i^{out} in (4) are related to the charge/discharge efficiency of BES devices.

In addition to the constraints for the components in MGs mentioned above, when the power deviation could be eliminated within one MG (i.e., autonomous operation of such single MG is achieved), it is unnecessary to exchange energy with other MGs based on the energy management principle of EI [14]. Typically, if one of the cases in (5) and (6) is satisfied, MG_i has the ability to absorb its inside power fluctuations. Thus, the action to dispatch energy from other MGs for MG_i would be unwise and should be avoided.

$$Case\ 1: \begin{cases} SOC_i \geq L_i^{SOC}, \\ 0 \leq P_i^{UC} \leq U_i^{MT} + U_i^{DEG} + U_i^{BES}, \\ P_i^{ER} \geq 0, \end{cases} \tag{5}$$

and

$$Case\ 2: \begin{cases} SOC_i \leq U_i^{SOC}, \\ -U_i^{BES} \leq P_i^{UC} \leq 0, \\ P_i^{ER} \leq 0. \end{cases} \tag{6}$$

3. Problem formulation

In this section, several types of cost for the operation of EI system is introduced. Some related penalty functions are designed. After that, the optimal control problem for the considered EI system under the constraints is formulated.

Let us denote the state space and action space of the considered system as $\mathcal{S}$ and $\mathcal{A}$, respectively. At each time step t, the state variable $s(t) \in \mathcal{S}$ of the considered EI system consists of $P_i^{UC}, P_i^{BES}, P_i^{UB}, SOC_i, (i \in V)$ and t.

Let

$$\begin{aligned} s_{UC}(t) &= \left[P_1^{UC}(t), \ldots, P_i^{UC}(t), \ldots, P_N^{UC}(t)\right]^{'}, \\ s_{BES}(t) &= \left[P_1^{BES}(t), \ldots, P_i^{BES}(t), \ldots, P_N^{BES}(t)\right]^{'}, \\ s_{UB}(t) &= \left[P_1^{UB}(t), \ldots, P_i^{UB}(t), \ldots, P_N^{UB}(t)\right]^{'}, \\ s_{SOC}(t) &= \left[SOC_1(t), \ldots, SOC_i(t), \ldots, SOC_N(t)\right]^{'}. \end{aligned}$$

Denote

$$s(t) = \left[s_{UC}(t)^{'}, s_{BES}(t)^{'}, s_{UB}(t)^{'}, s_{SOC}(t)^{'}, t\right]^{'}. \tag{7}$$

The controllable components are the power of ERs, MTs and DEGs.

Let

$$\begin{aligned} a_{ER}(t) &= \left[P_{1,1}^{ER}(t), \ldots, P_{i,j}^{ER}(t), \ldots, P_{N,N}^{ER}(t)\right]^{'}, \\ a_{MT}(t) &= \left[P_1^{MT}(t), \ldots, P_i^{MT}(t), \ldots, P_N^{MT}(t)\right]^{'}, \\ a_{DEG}(t) &= \left[P_1^{DEG}(t), \ldots, P_i^{DEG}(t), \ldots, P_N^{DEG}(t)\right]^{'}, \end{aligned}$$

Then, the controller $a(t) \in \mathcal{A}$ can be formulated as

$$a(t) = \left[a_{ER}(t)^{'}, a_{MT}(t)^{'}, a_{DEG}(t)^{'}\right]^{'}. \tag{8}$$

The initial state at t_0 is denoted as s_0. At each time step t, the controller $a(t)$ is obtained from a control scheme $u(s(t), t) \in \mathcal{U}$ and the system state $s(t)$.

3.1. Cost function for the EI system

The operation of the EI system during time interval $t \in [0, T]$ is considered. Since the power estimations for PVs, WTGs and loads in this paper are discrete time series, the EI system is studied in a discretization fashion. Suppose that there are $M+1$ estimation data during $[0, T]$, the time range is then discretized to be $M+1$ time steps, i.e., $t_0, t_1, \ldots, t_M$. The length between every two time steps is set to be $\Delta t = t_{k+1} - t_k = T/(M+1), k = 0, 1, \ldots, M$.

Firstly, the cost for power transmission between MGs are considered. In real-world power systems, transmission loss always occurs due to the long-distance power transmission and electrical conversions in converters [45, 46]. Thus, the following relationships are established.

$$C_{i.j}^{ER} = C_{j,i}^{ER}, \quad i, j \in V,$$
$$C_{i,i}^{ER} = 0, \quad i \in V,$$

where $C_{i,j}^{ER}$ is the transmission loss coefficient for the power line between ER_i and ER_j. In the field of EI, the transmission loss can be measured with the power of the related ER, and the cost for energy transmission from time step t_k to t_{k+1} can be described by

$$\Delta J_{ER}(t_k) = \frac{1}{2} \sum_{(i,j)\in E} C_{i.j}^{ER} \big| P_{i,j}^{ER}(t_k) \big| \Delta t.$$

Noted that since the same transmission loss is calculated twice in the summation above, $\frac{1}{2}$ is used to modify the result. Let us denote J_{ER} as the total cost for ERs within $[0, T]$. Then, we have

$$J_{ER} = \sum_{k=0}^{M} \Delta J_{ER}(t_k). \tag{9}$$

Apart from the cost of power transmission, the remarkable operation cost brought by MTs and DEGs are also worth considering. During the normal operation of the EI system, output power of MTs and DEGs can be properly controlled to meet the power demand. If irrational control schemes are applied, for example, in any MG, if power generation by PV units and WTGs is already enough for power consumption, and MTs and DEGs are still producing power consistently, then such status would significantly increase the operation cost of the EI system. Here, we assume that such cost is proportion to the output power of MTs and DEGs. From time step t_k to t_{k+1}, the cost of generators can be measured by

$$\Delta J_G(t_k) = \sum_{i\in V} (C_i^{MT} P_i^{MT}(t_k) + C_i^{DEG} P_i^{DEG}(t_k)) \Delta t,$$

where C_i^{MT} and C_i^{DEG} are constant coefficients for MTs and DEGs in MG_i, and they are related to the price of fuels and other concerned factors. The total cost of generators in the considered time period is given in (10).

$$J_G = \sum_{k=0}^{M} \Delta J_G(t_k). \tag{10}$$

According to [47, 48], the lifetime of BES devices could be measured by the Puekert lifetime energy throughput (PLET) model. The battery lifetime energy throughput c^{PLET} in the PLET model is defined as

$$c^{PLET} \triangleq (1-s)^{k_P} n,$$

where s is SOC of BES devices; k_P is the Peukert lifetime constant and it is usually within the range $[1.1, 1.3]$; n is the total number of battery cycles. As is introduced in [47], for any specified lower bound for SOC in the charge/discharge cycle of BES devices, the total c^{PLET}, which is denoted as C^{PLET}, for given BES devices is nearly constant. So, it can be used as a criteria for the lifetime of BES devices. Since k_P is close to 1, approximation formula for the reduction of c^{PLET} during a charge/discharge process of BES devices is derived based on [47] as follows:

$$\Delta c^{PLET} = \left(\sum_i \Delta s_i \right)^{k_P} \approx \sum_i {\Delta s_i}^{k_P},$$

where Δs_i is the SOC change in a short time period. Thus, the reduction for c^{PLET} of BES devices at time t can be approximated with

$$\Delta c^{PLET}(t) = |\Delta s(t)|^{k_P}.$$

We denote the total Peukert lifetime throughput and Peukert lifetime constant of the BES devices in MG_i as C_i^{PLET} and k_i^P, respectively. The loss of lifetime of BES devices, denoted as ΔL_i, in MG_i during the considered time period is

formulated in (11).

$$\begin{aligned}\Delta L_i &= \frac{\Delta c_i^{PLET}}{C_i^{PLET}} \\ &= \sum_{k=1}^{M} \Delta c_i^{PLET}(t_k)/C_i^{PLET} \\ &= \sum_{k=1}^{M} |SOC_i(t_k) - SOC_i(t_{k-1})|^{k_i^P}/C_i^{PLET}, \end{aligned} \tag{11}$$

To obtain the cost function for BES devices, the dynamics of the SOC in MG_i is rewritten in the discretization form in (12).

$$SOC_i(t_k) = SOC_i(t_{k-1}) - \eta_i P_i^{BES}(t_{k-1})\Delta t/Q_i. \tag{12}$$

With (11) and (12), the cost for BES devices from t_{k-1} to t_k is formulated in (13).

$$\begin{aligned}\Delta J_{BES}(t_k) &= \sum_{i\in V} |SOC_i(t_k) - SOC_i(t_{k-1})|^{k_i^P}/C_i^{PLET} \\ &= \frac{{\eta_i}^{k_i^P}}{{Q_i}^{k_i^P} C_i^{PLET}} \sum_{i\in V} \left(\left|P_i^{BES}(t_{k-1})\right|\Delta t\right)^{k_i^P}. \end{aligned} \tag{13}$$

So, the objective function for BES lifetime extension can be calculated from (14).

$$J_{BES} = \sum_{k=0}^{M} \Delta J_{BES}(t_k). \tag{14}$$

3.2. Penalty functions

In order that the constraints considered for the system in Section 2 hold during the operation of the EI system, a series of penalty functions are required to be formulated as follows.

Given the power of PVs, WTGs, DEGs, MTs, ERs, BES devices, loads, and SOC of BES devices at time step t_k, penalty functions are used to represent the constraints for the EI system. When all of the constraints hold, all of the penalty function are set to be zero. Whereas when there is one or more constraints been violated, the corresponding penalty functions will be assigned with a positive

value. To simplify the formulas, the characteristic function is employed. The characteristic function $\mathbb{I}(x)$ is defined as

$$\mathbb{I}(x) \triangleq \begin{cases} 1, & if\ x\ is\ true, \\ 0, & if\ x\ is\ false, \end{cases}$$

where x is a logical expression.

For the constraints of ERs, the penalty function $\phi^{ER}(t_k)$ is formulated as

$$\phi^{ER}(t_k) = \frac{1}{2} \sum_{(i,j)\in E} \Delta_{i,j}^{ER}(t_k) \mathbb{I}(\Delta_{i,j}^{ER}(t_k) \geq 0),$$

where

$$\Delta_{i,j}^{ER}(t_k) = \left| P_{i,j}^{ER}(t_k) \right| - U_{i,j}^{ER}.$$

For the constraints of MTs and DEGs, we set two penalty functions $\phi^{G}(t_k)$ and $\phi^{dG}(t_k)$. Here, $\phi^{G}(t_k)$ is used to restrict the output power of MTs and DEGs, and $\phi^{dG}(t_k)$ is used to avoid the over-control of MTs and DEGs. Let

$$\begin{aligned} \phi^{G}(t_k) =& \sum_{i\in V} \Delta_i^{MT}(t_k) \mathbb{I}(\Delta_i^{MT}(t_k) \geq 0) + \Delta_i^{DEG}(t_k) \mathbb{I}(\Delta_i^{DEG}(t_k) \geq 0), \\ \phi^{dG}(t_k) =& \sum_{i\in V} \Delta P_i^{MT}(t_k) \mathbb{I}(\Delta P_i^{MT}(t_k) \geq 0) + \Delta P_i^{DEG}(t_k) \mathbb{I}(\Delta P_i^{DEG}(t_k) \geq 0), \end{aligned}$$

in which

$$\begin{aligned} \Delta_i^{MT}(t_k) =& P_i^{MT}(t_k) - U_i^{MT}, \\ \Delta_i^{DEG}(t_k) =& P_i^{DEG}(t_k) - U_i^{DEG}, \\ \Delta P_i^{MT}(t_k) =& \left| P_i^{MT}(t_k) - P_i^{MT}(t_{k-1}) \right| - V_i^{MT}, \\ \Delta P_i^{DEG}(t_k) =& \left| P_i^{DEG}(t_k) - P_i^{DEG}(t_{k-1}) \right| - V_i^{DEG}, \end{aligned}$$

where V_i^{MT} and V_i^{DEG} are the upper bounds for the output power change of MTs and DEGs between two adjacent time steps, respectively. With such penalty for the power fluctuations of MTs and DEGs, the policies that may lead to over-control shall not be regarded as optimal.

For the constraints of BES devices, let us set penalty functions

$$\begin{aligned}\phi^{BES}(t_k) =& \sum_{i\in V} \Delta_i^{BES}(t_k)\mathbb{I}(\Delta_i^{BES}(t_k) \geq 0) + \Delta_i^{UB}(t_k),\\ \phi_{t_k}^{SOC} =& \sum_{i\in V} \mathbb{I}(SOC_i \leq L_i^{SOC}) + \mathbb{I}(SOC_i \geq U_i^{SOC}),\end{aligned}$$

where

$$\begin{aligned}\Delta_i^{BES}(t_k) =& \left|P_i^{BES}(t_k)\right| - U_i^{BES},\\ \Delta_i^{UB}(t_k) =& \left|P_i^{UB}(t_k)\right|.\end{aligned}$$

For the basic energy management principle of EI introduced in Section 1, let us set penalty functions

$$\phi^{EI}(t_k) = \sum_{i\in V} -P_i^{ER}\mathbb{I}(P_i^{ER} \leq 0)C_1 + P_i^{ER}\mathbb{I}(P_i^{ER} \geq 0)C_2,$$

where

$$\begin{aligned}C_1 =& \mathbb{I}(SOC_i \geq L_i^{SOC})\mathbb{I}(0 \leq P_i^{UC}(t_k) \leq U_i^{MT} + U_i^{DEG} + U_i^{BES}),\\ C_2 =& \mathbb{I}(SOC_i \leq UB_i^{SOC})\mathbb{I}(P_i^{UC}(t_k) \leq 0)\mathbb{I}(P_i^{UC}(t_k) + U_i^{BES} \geq 0).\end{aligned}$$

For the simplicity of the problem, all of the above penalty functions are summed with different weight factors, and the combined penalty function for the EI system at time step t_k is

$$\begin{aligned}\phi(t_k) =& \varepsilon_{ER}\phi^{ER}(t_k) + \varepsilon_G\phi^G(t_k) + \varepsilon_{dG}\phi^{dG}(t_k) + \varepsilon_{BES}\phi^{BES}(t_k)\\ &+ \varepsilon_{SOC}\phi^{SOC}(t_k) + \varepsilon_{EI}\phi^{EI}(t_k),\end{aligned}$$

where $\varepsilon_{ER}, \varepsilon_G, \varepsilon_{dG}, \varepsilon_{BES}, \varepsilon_{SOC}, \varepsilon_{EI}$ are weight factors for different penalty functions.

The penalty function for the considered time period is then calculated as

$$\Phi = \sum_{k=0}^{M} \phi(t_k)\Delta t. \tag{15}$$

Any control scheme that causes the violation of these constraints will lead to a nonzero value of (15). In other words, if the penalty function during the considered period is minimized, then no constraint is violated.

3.3. Optimal control problem with constrains

For the energy management issue of the considered EI system, all of the costs derived in (9), (10) and (14) need to be taken into consideration. To achieve the trade-off of these costs, the cost function to be minimized is formulated as their weighted sum, given as follows,

$$J = \alpha_{ER} J_{ER} + \alpha_G J_G + \alpha_{BES} J_{BES}, \tag{16}$$

where scalars α_{ER}, α_G and α_{BES} are the weight coefficients. By properly adjusting the weight coefficients in (16), different optimal objectives can be achieved. For example, if we set α_{ER} to be significantly larger than the rest two coefficients, the optimal control scheme would emphasize to reduce the amount of energy exchange among MGs. If J_{BES} is emphasized, the optimal control policy would rely more on ERs to absorb power deviations in the considered EI system.

Our goal is to find the optimal control scheme $u^*(s(t), t)$, such that the sum of cost function (16) and the penalty function (15) is minimized. In this sense, the optimal control problem can be rewritten as (time t omitted)

$$\begin{aligned} \min_{u \in \mathcal{U}} \quad & \mathbb{E}[J + \Phi], \\ \text{subject to} \quad & s(t_0) = s_0, \end{aligned} \tag{17}$$

where $\mathbb{E}$ is the mathematical expectation. Due to the stochastic character of P_i^{UC}, both J and ϕ are stochastic processes. So, the expectation operator is used here.

4. Solution to the optimal control problem

Instead of solving the Bellman equation directly, there are several solvable methods for the HJB/Bellman equation; see, e.g., [35, 36]. They are able to deal with systems similar as (17). However, almost all of these solutions use "grid based" methods [35, 36] which means that they rely on the discretization of action space and state space. As a result, these methods suffer from the

"curse of dimensionality" when the dimension of action space and state space becomes larger [37]. For the considered EI system, the dimension of action space is $2|V|+|E|$ and the dimension of the state space is $4|V|+1$. In real scenarios of EI, since there may exist a number of MGs, it is obvious that these grid based approaches are not applicable for EI systems.

Meanwhile, in this paper, a set of constraints are set for the considered EI system. These constraints make it even harder to obtain solutions with conventional methods. Fortunately, with the help of deep reinforcement learning approach, it is possible to obtain practical solutions for our problem. Noted that not all reinforcement learning techniques can be applied to our considered optimal control problem. The value based approaches will suffer from the curse of dimensionality, due to the continuous action space. In this paper, the cutting-edge reinforcement learning technique named asynchronous actor-critic agents (A3C) [38] is employed to find solutions to (17).

4.1. Converting optimal control problem to reinforcement learning problem

Here, we convert the optimal control problem into a suitable form for the reinforcement learning issue. In a reinforcement learning problem, there are an agent and an environment. The agent interacts with the environment based on certain control policy and the state observed from the environment. At each time, a reward is provided to the agent as the feedback for the action taken by the agent. By exploring the action space $\mathcal{A}$, the agent learns the optimal control policy that maximizes the total reward.

In this paper, the EI system is the environment for the agent. The agent is assumed to control the power of ERs, MTs, and DEGs in MGs. At time step $t_k \in [0, T]$, the environment provides the system state to the agent. The agent generates action $a(t_k)$ based on its control policy π and the observed system state $s(t_k)$. According to the EI system modelling, the sum of the uncontrollable components P_i^{UC} in MG_i is a stochastic process. Other components in the state variable are deterministic variables. Since the scalar Weiner process P_i^E in P_i^{UC} has Markov property, the transition probability from $s(t_{k-1})$ to $s(t_k)$ is only

related to the action $a(t_{k-1})$ and $s(t_{k-1})$, as is described in (18).

$$P_a(s, s') = \mathbb{P}\{s(t_k) = s' | s(t_{k-1}) = s, a(t_{k-1}) = a)\}, \quad s, s' \in \mathcal{S}, \quad a \in \mathcal{A}. \quad (18)$$

From t_k to t_M, the total reward R_{t_k} is

$$R_{t_k} = \sum_{i=0}^{M} \gamma^i r(t_{i+k}), \quad (19)$$

where $\gamma \in [0, 1]$ is the attenuation coefficient; $r(t_{i+k})$ is the reward for the state transition from $s(t_{k-1})$ to $s(t_k)$ with action $a(t_k)$, and

$$r(t_{i+k}) = r_{a(t_k)}(s(t_k), s(t_{k-1})).$$

Given a policy π, the value function of for state s at time step t_k is

$$V^{\pi}(s(t_k)) = \mathbb{E}[R_{t_k} | s(t_k) = s].$$

The target for the agent is to find the optimal control policy π^* that maximizes $V^{\pi^*}(s_0, t_0)$.

In this paper, the attenuation coefficient γ is set to be 1, such that the rewards can directly correspond to the target $J + \Phi$ in (17). The design for the reward at each time step is demonstrated as follows.

Based on the costs and penalty functions formulated in Section 3, the reward at time step t_k is derived as

$$r_{t_k} = -\alpha_{ER} \Delta J_{ER}(t_k) - \alpha_G \Delta J_G(t_k) - \alpha_{BES} \Delta J_{BES}(t_k) - \phi(t_k) \Delta t.$$

Thus, the following relationship is established,

$$V^{\pi}(s(t_0)) = -\mathbb{E}[J + \phi].$$

The optimal controller u^* for (17) is equivalent to the optimal policy π^* that maximizes $V^{\pi}(s(t_0))$.

Now, the discrete time EI system is described with a Markov decision process [49] $(\mathcal{S}, \mathcal{A}, P_{\cdot}(\cdot, \cdot), r_{\cdot}(\cdot, \cdot), \gamma)$. It can be solved with the reinforcement learning approaches [50].

4.2. A3C algorithm and network structure

In the A3C algorithm, the actor-critic architecture is applied. The value function $V^{\pi}(s(t_k))$ is estimated with a neural network "critic". The control policy π is approximated with another neural network "actor". To capture the potential temporal features, recurrent neural network (RNN) [51] is constructed as the first layer of the whole network. As is shown in Figure 3, the normalized state $s(t)$ is fed as the input of the RNN layer, and the output of this layer is assigned to the critic and actor networks simultaneously. The critic network consists of two full connection layers. The output is a scalar which is denoted as $v(s(t);\theta_c)$. Similar as the critic network, the actor network has two full connection layers. For better exploration performance, Gaussian policy [52] is used to obtain the controller at each time. Thus, there are two outputs of the actor network. One is the mean value of the action $\mu(s(t);\theta_a)$, and the other is the standard variance of the action $\sigma(s(t);\theta_a)$. The action $a(t)$ is sampled from the normal distribution $\mathcal{N}(\mu(s(t)), \sigma^2(s(t)); \theta_a)$. Here, θ_c and θ_a are parameters of the two neural networks.

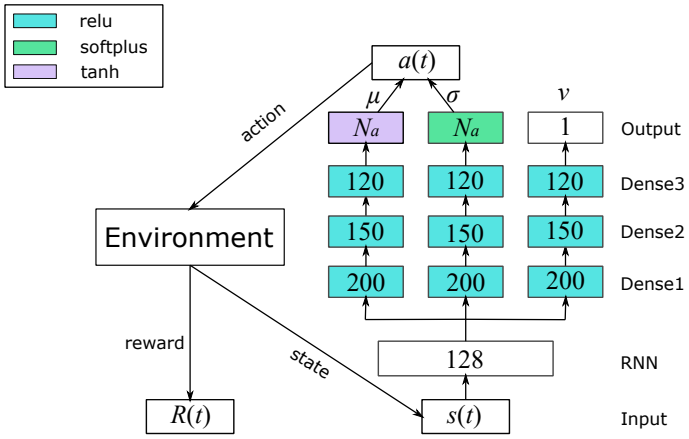

Figure 3: network structure

According to [38], the gradients for the critic and actor networks are calcu-

lated with

$$\frac{\partial}{\partial \theta_c}(R_t|s(t) - v(s(t);\theta_c))^2,$$

and

$$\nabla_{\theta_a} \log \mathbb{P}\{a(t)|s(t);\theta_a\}(R_t|s(t) - v(s(t);\theta_c)).$$

During the training, there are multiple threads running asynchronously. In each thread, the network in Figure 3 is constructed and used to generate the controller at each time step. The states of the environments in different threads are updated independently. Meanwhile, a global network is maintained. Once a thread collects a series data of n steps, the parameters of the global network are updated with these data. After that, the parameters of the network in the same thread will sync with the global network. By training in this way, the correlation between the training data is eliminated. Thus, the "replay" technique is unnecessary and the training process is more efficient.

By applying the A3C algorithm in the training of the neural network designed in Figure 3, the intelligent controller for the EI system can be contained finally. Given an observation of the EI system, the network will generate corresponding controller to achieve an intelligent operation.

5. Simulation

In this section, the effectiveness of the proposed energy management strategies for EI system is evaluated. Although the sub-optimal solutions to our optimal control problem could be found by some heuristic algorithms, e.g., particle swarm optimization (PSO) [39], genetic algorithm (GA) [40], simulate anneal arithmetic (SAA) [41], etc., due to the large search space, it will be difficult to find an appropriate solution to the energy management problem with these conventional methods. Besides, in this paper, the constraints for real EI system is formulated as penalty functions, which will essentially lead to the failure of these heuristic algorithms. Thus, only the feasibility of the proposed control method is evaluated in this section.

Without of loss of generality, the numerical simulation is carried out on a network consists of four MGs and four ERs. The topology of the investigated system is shown in Figure 4 where MG_1 is interconnected with MG_2; MG_2 is interconnected with MG_1, MG_3, and MG_4; MG_3 and MG_4 are interconnected with each other. As is mentioned in Section 2, all of these MGs are assumed to consist of similar components. In case of equipment damage, we assume that MTs and DEGs in MG_4 are out of order. Thus, the realization of power balance in MG_4 would rely heavily on power exchange via ERs. According to [21], such EI topology can be extended to a generalized EI scenario without essential difficulty.

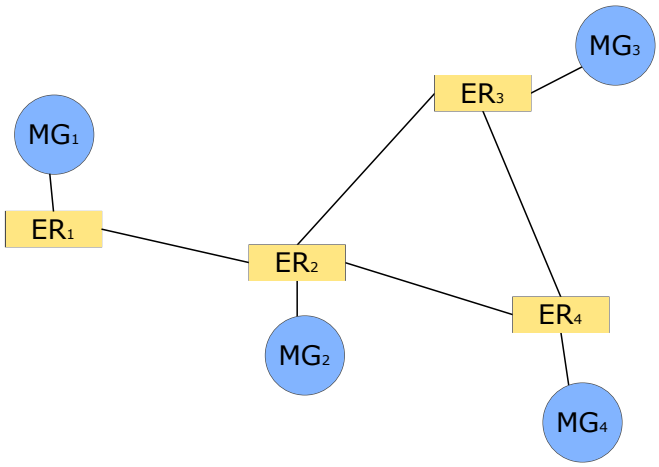

Figure 4: simulation-EI-topology

The simulation time period is set to be one day, e.g., $t \in [0, 24]$ (time unit *hour* omitted). The data used as the power forecast results for PVs, WTGs and loads are generated from [33]. The parameters for the simulation are given in Table 1.

By training the neural network with A3C algorithm [38], the intelligent control scheme for the EI network considered in this section is obtained. The curves for power flow of ERs are plotted in Figure 5. The detailed power dynamics of MG_1, MG_2, MG_3 and MG_4 are presented in Figure 6, Figure 7, Figure 8 and

Parameters	Value	Parameters	Value
$U_{i,j}^{ER}, i,j=1,2,3,4$	2000(kW)	$U_i^{MT}, i=1,2,3,4$	900(kW)
$U_i^{DEG}, i=1,2,3,4$	800 (kW)	$U_i^{BES}, i=1,2,3,4$	600(kW)
$V_i^{MT}, i=1,2,3,4$	20(kW)	$V_i^{DEG}, i=1,2,3,4$	30(kW)
$L_i^{SOC}, i=1,2,3,4$	0.2	$U_i^{SOC}, i=1,2,3,4$	0.8
$\eta_i^{in}, i=1,2,3,4$	0.96	$\eta_i^{out}, i=1,2,3,4$	0.97
$C_i^{MT}, i=1,2,3,4$	0.004	$C_i^{DEG}, i=1,2,3,4$	0.005
$C_{1,2}^{ER}$	0.24	$C_{2,3}^{ER}$	0.23
$C_{2,4}^{ER}$	0.31	$C_{3,4}^{ER}$	0.15
C_1^{PLET}	23	C_2^{PLET}	23
C_3^{PLET}	23	C_4^{PLET}	23
Q_1	80(kWh)	Q_2	40(kWh)
Q_3	55(kWh)	Q_4	50(kWh)
k_i^P	1.075	α_{ER}	3.6
α_G	0.7	α_{BES}	0.1
ε_{ER}	3.0	ε_G	3.0
ε_{dG}	3.0	ε_{BES}	3.0
ε_{SOC}	3.0	ε_{EI}	1.0

Table 1: Parameters

Figure 9, respectively.

From Figure 6, the SOC of BES devices is properly maintained within the lower bound and upper bound set in Table 1. It is notable that within the time period $[0, 12]$, MG_1 is able to achieve power balance without exchanging energy with the energy routing network. During time period $[12, 18]$, the output power of PVs grows rapidly with the increasing solar irradiation. In order that the SOC of BES devices does not exceed the upper bound U_1^{SOC}, MG_1 transmits the redundant energy to the energy routing network. Thus, there is a trough in the power curve of P_1^{ER} in such period.

In MG_2, it is assumed that the local loads require a plenty of electric power.

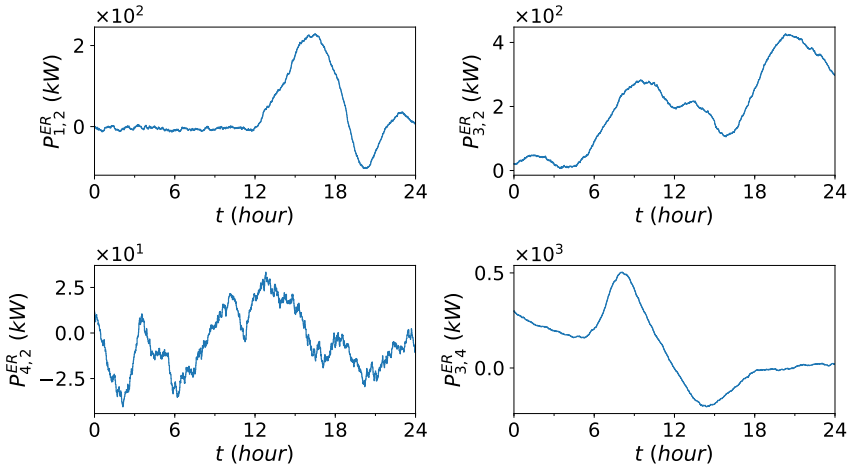

Figure 5: Power dynamics of ERs

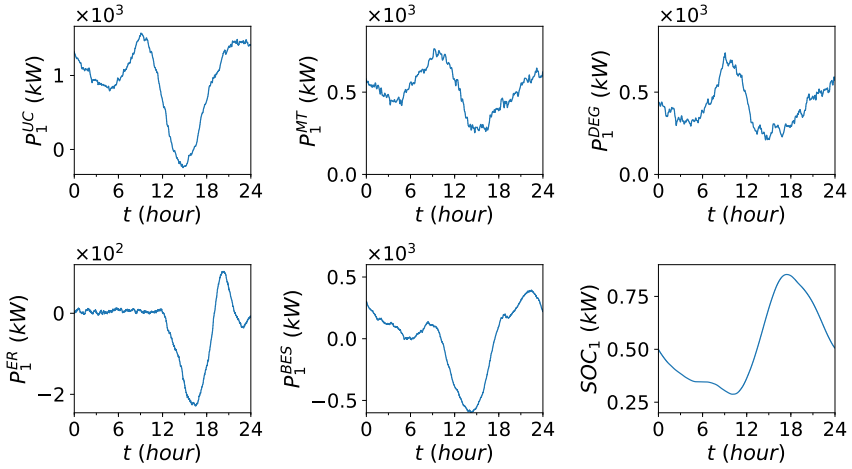

Figure 6: Power dynamics of MG_1

To compensate such power consumption, the output power of MTs and DEGs shall be controlled at a high level, as is shown in Figure 7. In order to protect the BES devices as well as to consume energy shared by other MGs, energy is transmitted to MG_2 consistently via the energy routing network. According to Figure 8, abundant power is generated by PVs and WTGs in MG_3. Since the

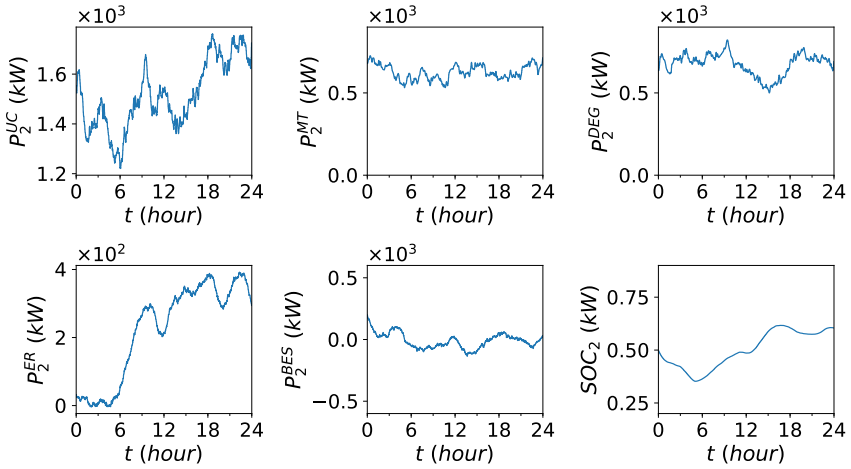

Figure 7: Power dynamics of MG_2

capacity of BES devices is limited, MG_3 would share more power to the energy routing network, as is presented in Figure 5.

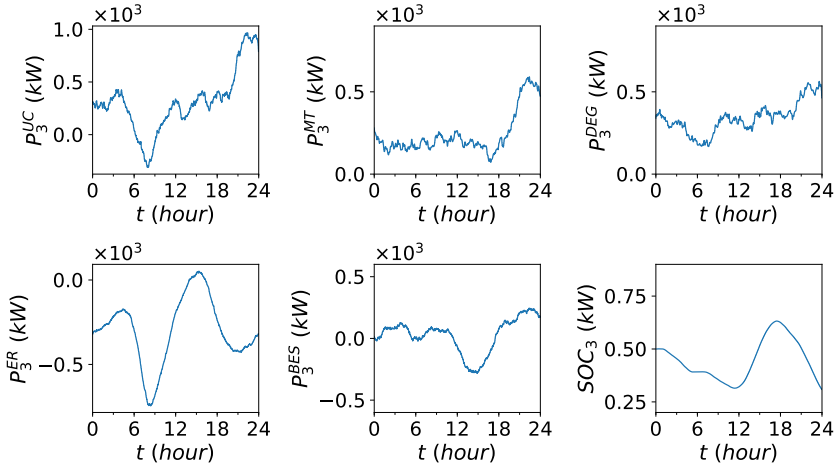

Figure 8: Power dynamics of MG_3

The dynamics of MG_4 is illustrated in Figure 9 where we find that ERs play an important role for MG_4 operation. In the considered time period, although the MTs and DEGs in MG_4 are not able to function normally, with the help of

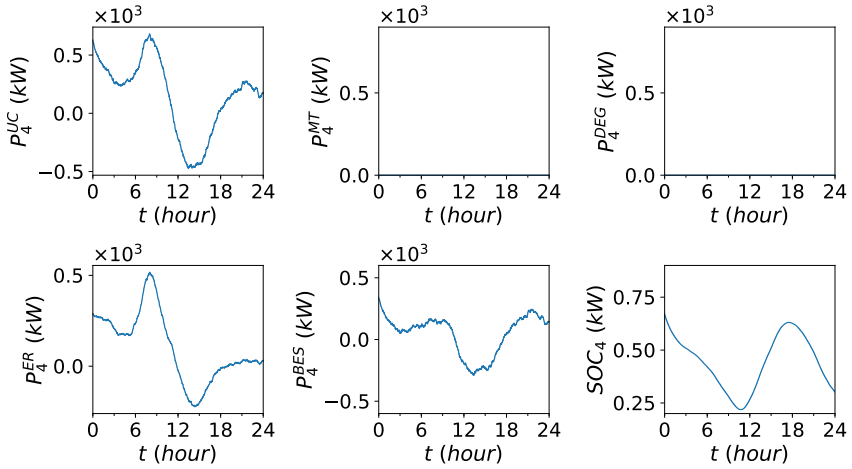

Figure 9: Power dynamics of MG_4

the ER network, the power balance is still achieved in MG_4, and the SOC of BES devices has been kept in a proper range.

Based on the simulation result, the feasibility and effectiveness of the obtained controller is evaluated. The advantages of the EI system in which MGs in different areas are interconnected via ERs are demonstrated. Since the situations in different MGs are diverse, the energy routing network can fully utilize the available resources and capacities in the EI system and provide more reliable power supply.

6. Conclusion

In this paper, the energy management issue for a generalized EI system is investigated. The deep reinforcement learning approach is applied to solve such control problem. The simulation results shows the effectiveness of the proposed method. In the future, it is also important to develop distributed control schemes for EI scenarios, such that the energy management strategies for the whole system would become more flexible and robust.

7. Acknowledgement

This work was supported in part by National Natural Science Foundation of China (grant No. 61472200) and Beijing Municipal Science & Technology Commission (grant No. Z161100000416004).

References

[1] Bilgen S, Kaygusuz K, Sari A. Renewable energy for a clean and sustainable future. Energy Source 2004;26:1119-29.

[2] Mathiesen BV, Lund H, Connolly D, Wenzel H, stergaard PA, Mller B, et al. Smart energy systems for coherent 100% renewable energy and transport solutions. Appl Energy 2015;145:139-154.

[3] Vlachogiannis JG. Probabilistic constrained load flow considering integration of wind power generation and electric vehicles. IEEE Trans Power Syst 2009;24:1808-17.

[4] Bevrani H, Feizi MR, Ataee S. Robust frequency control in an islanded microgrid: H_∞ and μ-synthesis approaches. IEEE Trans Smart Grid, 2016;7:706-17.

[5] Venkataramanan G, Marnay C. A larger role for microgrids. IEEE Power Energy Mag 2008;6:7882.

[6] Elsayed AT, Mohamed AA, Mohammed OA. DC microgrids and distribution systems: An overview. Elect Power Syst Res 2015;119:407-17.

[7] Kou P, Liang D, Gao L. Distributed EMPC of multiple microgrids for coordinated stochastic energy management. Appl Energy 2017;185:939-52.

[8] Korkas CD, Baldi S, Michailidis I, Kosmatopoulos E. Occupancy-based demand response and thermal comfort optimization in microgrids with renewable energy sources and energy storage. Appl Energy 2016;163:93-104.

[9] Hua H, Qin Y, Cao J. A class of optimal and robust controller design for islanded microgrid. In: IEEE 7th international conference on power and energy systems. Toronto, Canada; 2017. p. 111-6.

[10] Hua H, Qin Y, Cao J, Wang W, Zhou Q, Jin Y, et al. Stochastic optimal and robust control scheme for islanded AC microgrid. In: IEEE international conference on probabilistic methods applied to power systems. Boise, Idaho, US; 2018. p. 78-84.

[11] Rifkin J. The Third Industrial Revolution: How Lateral Power is Transforming Energy, the Economy, and the World. Palgrave Macmillan, New York, US; 2013. p. 31-46.

[12] Dong Z, Zhao J, Wen F, Xue Y. From smart grid to energy internet: basic concept and research framework. Automat Elec Power Syst 2014;38:1-11.

[13] Tsoukalas LH, Gao R. From smart grids to an energy Internet - assumptions, architectures and requirements. Smart Grid & Renew Energy 2009;1:18-22.

[14] Cao J, Yang M. Energy Internet - towards smart grid 2.0. In: 4th international conference on networking & distributed computing. Los Angeles, USA; 2013. p. 105-10.

[15] Han X, Yang F, Bai C, Xie G, Ren G, Hua H, Cao J. An open energy routing network for low-voltage distribution power grid. In: 1st IEEE international conference on energy Internet. Beijing, China; 2017. p. 320-5.

[16] Xu Y, Zhang J, Wang W, Juneja A, Bhattacharya S. Energy router: architectures and functionalities toward energy internet. In: 2011 IEEE international conference on smart grid communications. Brussels, Belgium; 2011. p. 31-6.

[17] Ma Y, Wang X, Zhou X, Gao Z. An overview of energy routers. In: 29th Chinese control and decision conference. Chongqing, China; 2017. p. 4104-8.

[18] Geidl M, Koeppel G, Favre-Perrod P, Klokl B. Energy hubs for the futures. IEEE Power & Energy Mag 2007;5:24-30.

[19] Boyd J. An internet-inspired electricity grid. IEEE Spectr 2013;50:12-4.

[20] Hua H, Cao J, Yang G, Ren G. Voltage control for uncertain stochastic nonlinear system with application to energy Internet: non-fragile robust H approach. J Math Anal Appl 2018;463:93-110.

[21] Hua H, Qin Y, Cao J. Coordinated frequency control for multiple microgrids in energy Internet: a stochastic H approach. In: 2018 IEEE PES Innovative Smart Grid Technologies Asia. Singapore; 2018. p. 247-53.

[22] Sun Q, Han R, Zhang H, Zhou J, Guerrero JM, A multiagent-based consensus algorithm for distributed coordinated control of distributed generators in the energy internet. IEEE Trans. Smart Grid 2015;6:3006-19.

[23] Wang R, Wu J, Qian Z, Lin Z, A graph theory based energy routing algorithm in energy local area network, IEEE Trans Ind Inform 2017;13:3275-85.

[24] Vachirasricirikul S, Ngamroo I. Robust controller design of microturbine and electrolyzer for frequency stabilization in a microgrid system with plug-in hybrid electric vehicles. Elect Power Energy Syst 2012;43:804-11.

[25] Vachirasricirikul S, Ngamroo I. Robust controller design of heat pump and plug-in hybrid electric vehicle for frequency control in a smart microgrid based on specified-structure mixed H_2/H_∞ control technique. Appl Energy 2011;88:3860-8.

[26] Mao X, Stochastic Differential Equations and Applications, Second Edition. Horwood Publishing, Chichester, UK, 2007.

[27] Odun-Ayo T, Crow ML. Structure-preserved power system transient stability using stochastic energy functions. IEEE Trans Power Syst 2012;27:1450-8.

[28] Marino DL, Amarasinghe K, Manic M. Building energy load forecasting using deep neural networks. In: 42nd annual conference of the IEEE industrial electronics society. Florence, Italy; 2016. p. 7046-51.

[29] Zhu H, Li X, Sun Q, Nie L, Yao J, Zhao G. A power prediction method for photovoltaic power plant based on wavelet decomposition and artificial neural networks. Energies 2015;9:1-15.

[30] Zhang W, Hua H, Cao J. Short term load forecasting based on IGSA-ELM algorithm. In: 1st IEEE international conference on energy Internet. Beijing, China; 2017. p. 296-301.

[31] Claessens BJ, 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 2016;99:1-11.

[32] Yang G, Cao J, Hua H, Zhou Z. Deep learning-based distributed optimal control for wide area energy Internet. In: 2nd IEEE international conference on energy Internet. Beijing, China; 2018. p. 292-7.

[33] "Dataport," Pecan Street Inc., https://dataport.cloud/.

[34] Festa A, Guglielmi R, Hermosilla C, Picarelli A, Sahu S, Sassi A, Silva FJ. HamiltonJacobiBellman equations. In: Optimal control: novel directions and applications. Springer; 2017. p. 127-261.

[35] Szpiro A, Dupuis P. Second order numerical methods for first order Hamilton-Jacobi equations. SIAM J Numerical Anal 2002;40:1136-83.

[36] Falcone M, Ferretti R. Convergence analysis for a class of high-order semi-Lagrangian advection schemes. SIAM J Numerical Anal 1998; 35:909-40.

[37] McEneaney WM, Deshpande A, Gaubert S. Curse-of-complexity attenuation in the curse-of-dimensionality-free method for HJB PDEs. In: American control conference. Seattle, US; 2008. p. 4684-4690.

[38] Mnih V, Badia AP, Mirza M, Graves A, Lillicrap T, Harley T, et al. Asynchronous methods for deep reinforcement learning. In: international conference on machine learning. New York, US; 2016. p. 1928-37.

[39] Couceiro M, Ghamisi P. Particle Swarm Optimization. Fractional Order Darwinian Particle Swarm Optimization. Springer International Publishing; 2016.

[40] Wang J, Ersoy OK, He M, Wang F. Multi-offspring genetic algorithm and its application to the traveling salesman problem. Appl Soft Computing 2016;43:415-23.

[41] Isakov SV, Zintchenko IN, Rnnow TF, Troyer M. Optimised simulated annealing for Ising spin glasses. Computer Physics Commun 2015;192:265-71.

[42] Varma RK, Rahman SA, Vanderheide T. New control of PV solar farm as STATCOM (PV-STATCOM) for increasing grid power transmission limits during night and day. IEEE Trans Power Del 2015;30:755-63.

[43] Alizadeh MI, Moghaddam MP, Amjady N, Siano P, Sheikh-El-Eslami, MK. Flexibility in future power systems with high renewable penetration: A review. Renew & Sustain Energy Reviews 2016;57:1186-93.

[44] Heymann B, Bonnans JF, Silva F, Jimenez G. A stochastic continuous time model for microgrid energy management. In: 2016 European control conference. Aalborg, Denmark; 2016. p. 2084-9.

[45] Expsito AG, Conejo AJ, Canizares C. Electric energy systems: analysis and operation. CRC press; Boca Raton, FL, US, 2016.

[46] Zhang Y, Rahbari-Asr N, Duan J, Chow MY. Day-ahead smart grid cooperative distributed energy scheduling with renewable and storage integration. IEEE Trans Sustain Energy 2016;7:1739-48.

[47] Tran D, Khambadkone AM. Energy management for lifetime extension of energy storage system in micro-grid applications. IEEE Trans Smart Grid 2013;4:1289-96.

[48] Lashway CR, Mohammed OA. Adaptive battery management and parameter estimation through physics-based modeling and experimental verification. IEEE Trans Transport Electrific 2016;2:454-64.

[49] Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science 2015;349:255-60.

[50] Turchetta M, Berkenkamp F, Krause A. Safe exploration in finite Markov decision processes with Gaussian processes. In: Advances in neural information processing systems. Barcelona, Spain; 2016. p. 4312-20.

[51] Sak H, Senior A, Beaufays F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. Computer Science 2014;338-42.

[52] Hachiya H, Peters J, Sugiyama M. Efficient sample reuse in EM-based policy search. In: Joint European conference on machine learning and knowledge discovery in databases. Springer, Berlin, Heidelberg; 2009. p. 469-84.