به نام خداوند بخشنده و مهربان

موضوع: ارائه سمینار

نام استاد: دکتر شمقدری

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2020.2976771, IEEE Transactions on Smart Grid

Model-Free Real-Time Autonomous Control for A Residential Multi-Energy System Using Deep Reinforcement Learning

Yujian Ye, *Member, IEEE,* Dawei Qiu, *Student Member, IEEE,* Xiaodong Wu, *Student Member, IEEE,* Goran Strbac, *Member, IEEE* and Jonathan Ward, *Member, IEEE*

***Abstract*—Multi-energy systems (MES) are attracting increasing attention driven by its potential to offer significant flexibility in future smart grids. At the residential level, the roll-out of smart meters and rapid deployment of smart energy devices call for autonomous multi-energy management systems which can exploit real-time information to optimally schedule the usage of different devices with the aim of minimizing end-users' energy costs. This**

η^{gb}	Gas-to-heat conversion efficiency of GB
B. Variables	
$P^{eg,ed}$	Power flow from the EG to ED (kW)
$P^{eg,hp}$	Power flow from the EG to EHP (kW)
$P^{eg,es}$	Power flow from the EG to EES (kW)
$P^{gg,gb}$	Gas input from the GG to GB (kW)

Activate Windows
Go to Settings to activat

ساختار سیستم چندحاملی مورد مطالعه

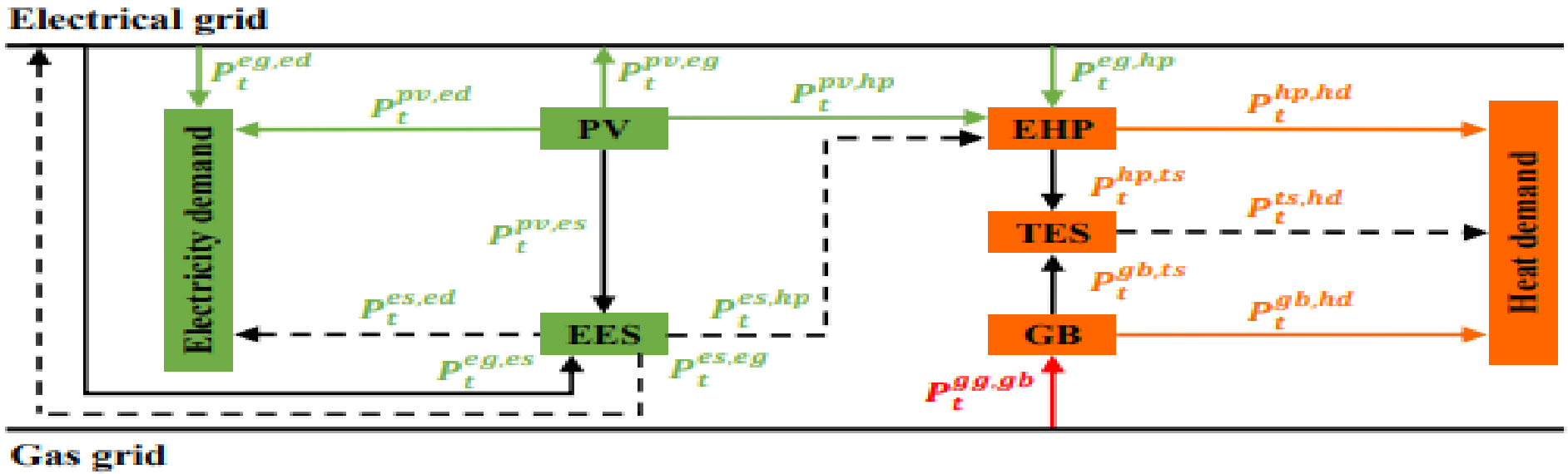

Fig. 1. Structure of the studied residential multi-energy system. The black solid and dashed arrows indicate mutually exclusive power flows.

بیان مساله در قالب MDP

- حالت ها:

$$s_t = [t, E_t, P_t^d, \lambda_t, P_t^{pv}] \qquad E_t = [E_t^{es}, E_t^{ts}]$$

$$P_t^d = [P_t^{ed}, P_t^{hd}], \ \lambda_t = [\lambda_t^{e-}, \lambda_t^{e+}, \lambda_t^g]$$

- اکشن ها:

$$a_t = [a_t^{esc/d}, a_t^{tsc/d}, a_t^{es}, a_t^{ts}, a_t^{gb,hd}, a_t^{gb,ts}, a_t^{pv,hp} a_t^{pv,es}, a_t^{pv,ed}, a_t^{es,ed}, a_t^{es,hp}]$$

Prioritized Deep Deterministic Policy Gradient (PDDPG)

روش PRIORITIZED EXPERIENCE REPLAY

- استفاده از بزرگی خطای TD به عنوان معیاری برای تصحیح تخمین مقادیر Q

تجربه با TD مثبت بزرگ تجربه خیلی موفق و تجربه با TD منفی بزرگ دارای خطای زیاد

اولویت دادن به تجربه ها باعث همگرایی سریع می شود

اجتناب ایجنت کارهای نامطلوب در برخی از حالت ها و افزایش کیفیت پالیسی یاد گرفته شده

روش PRIORITIZED EXPERIENCE REPLAY

- تعریف احتمال نمونه برداری تجربه i ام :

$$P_i = p_i^{\beta_1} / \sum_k p_k^{\beta_1}$$

$$p_i = 1/rank_i$$

$$|\delta_i| = \left| r_i + \gamma Q_{\theta'}\left(s_{i+1}, \mu_{\phi'}(s_{i+1})\right) - Q_\theta(s_i, a_i) \right| \quad (30)$$

- باعث افزودن بایاس - > معرفی وزن های نمونه

$$W_i = (N_{PR} P_i)^{-\beta_2} / \max_k W_k$$

روش PRIORITIZED EXPERIENCE REPLAY

$$L_\theta = \mathbb{E}\left[\left(r_t + \gamma Q_\theta\left(s_{t+1}, \mu_\phi(s_{t+1})\right) - Q_\theta(s_t, a_t)\right)^2\right]$$

$$L_\theta = \frac{1}{N}\sum_{i=1}^{N} W_i \delta_i^2$$

$$\nabla_\phi J(\mu_\phi) = \mathbb{E}_{s\sim\rho^\mu}\left[\nabla_a Q_\theta(s,a)|_{a=\mu_\phi(s)} \nabla_\phi \mu_\phi(s)\right].$$

$$\nabla_\phi J(\mu_\phi) = \frac{1}{N}\sum_{i=1}^{N} \nabla_a Q_\theta(s_i,a)|_{a=\mu_\phi(s_i)} \nabla_\phi \mu_\phi(s_i)$$

$$\theta \leftarrow \theta + \alpha^\theta \nabla_\theta L_\theta$$

$$\phi \leftarrow \phi + \alpha^\phi \nabla_\phi J(\mu_\phi)$$

آپدیت وزن های شبکه های Actor و Criticآنلاین

$$\theta' \leftarrow \tau\theta + (1-\tau)\theta'$$

$$\phi' \leftarrow \tau\phi + (1-\tau)\phi'$$

آپدیت وزن های شبکه های Actor و Critic هدف

Algorithm 1 Training procedure of PDDPG

Initialize the online critic and actor networks with random weights θ and ϕ, respectively.
Initialize the target critic and actor networks with weights $\theta' \leftarrow \theta$ and $\phi' \leftarrow \phi$, respectively.
for episode $m = 1 : M^{train}$ **do**
  Obtain the initial state s_1 from a random day in the training set.
  Initialize a random Gaussian exploration noise $\mathcal{N}_t$.
  for time step (i.e. hour) $t = 1 : T$ **do**
    Selects control action a_t using (28).
    Execute action a_t in the MES environment, observe reward r_t, and transit to the new state s_{t+1}.
    Store, in $\mathcal{PR}$, experience (s_t, a_t, r_t, s_{t+1}) and set $p_t = \max_{i<t} p_i$.
    for $i = 1 : N$ **do**
      Sample experience i with probability P_i in (29).
      Compute the absolute TD-error $|\delta_i|$ using (30).
      Compute the IS weights W_i using (31).
      Update the priority p_i according to $|\delta_i|$.
    end for
    Update the critic network using (32) and (34).
    Update the actor network using (33) and (35).
    Update the target networks using (36) and (37).
  end for
end for

Algorithm 2 PDDPG-based energy management strategy

Load the DNN's parameter ϕ^* of the online actor network μ_{ϕ^*} trained by Algorithm 1.
for test day = $1 : M^{test}$ **do**
  Obtain the initial state s_1 of the test day.
  for time step = $1 : T$ **do**
    Set the energy management action as $a_t = \mu_{\phi^*}(s_t)$.
    Execute action a_t in the MES environment, observe reward r_t, and transit to the new state s_{t+1}.
  end for
end for

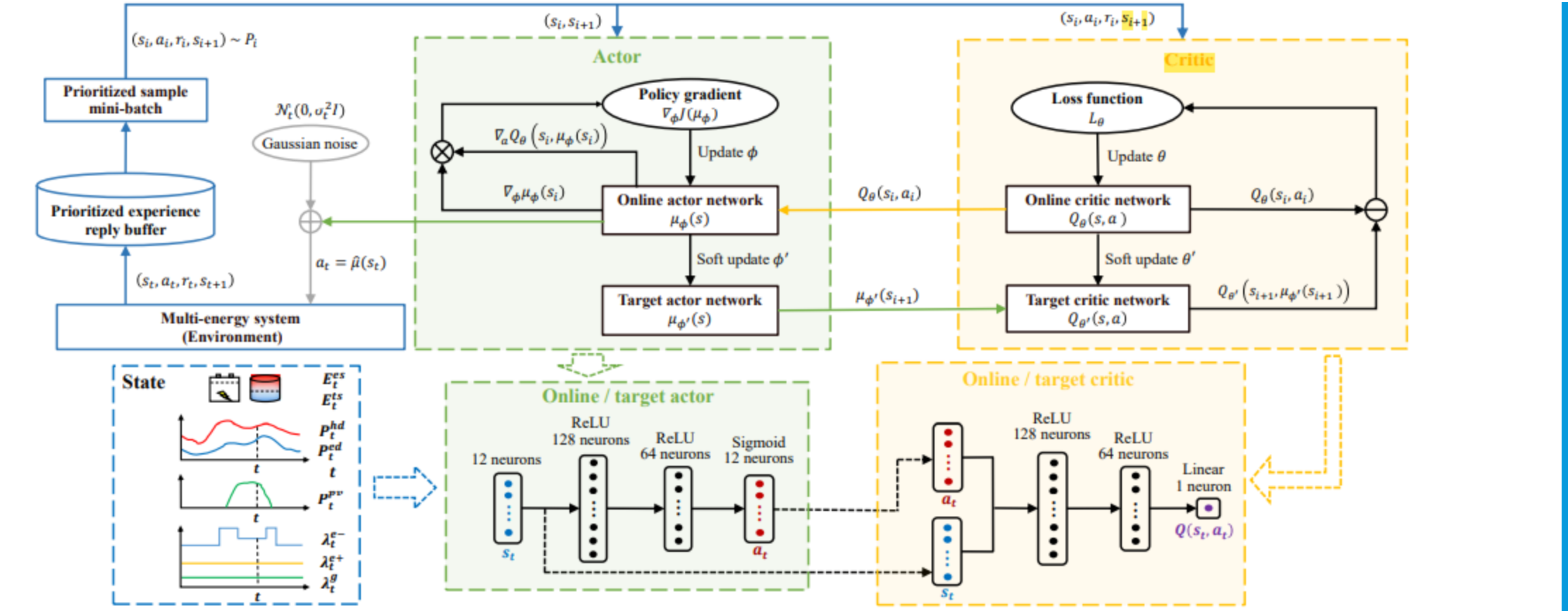

Fig. 3. Workflow of PDDPG.

تنظیم تقاضای برق در زمستان با روش PDDPG

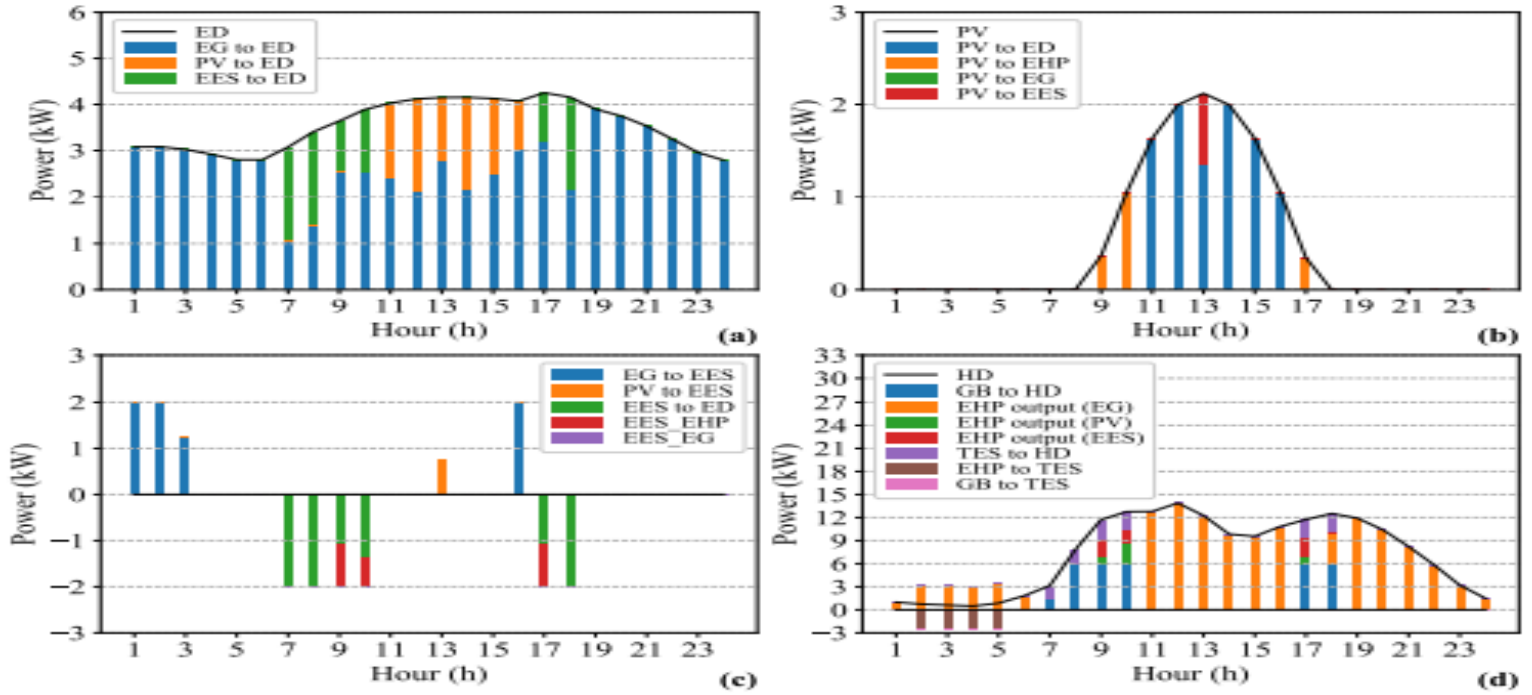

Fig. 6. (a) Balancing of ED, (b) usage of PV generation, (c) charging/discharging schedule of EES and (d) balancing of HD for the examined winter day under the PDDPG method.

مقایسه هزینه روزانه

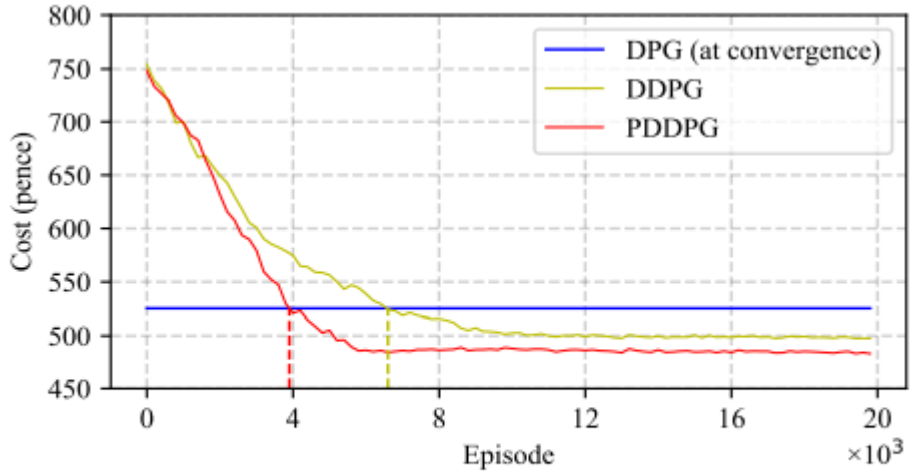

Fig. 9. Average daily cost over 10 different random seeds for the DDPG and PDDPG methods.

استفاده از مکانیزم تجربه دارای اولویت در PDDPG و نمونه برداری یکنواخت در DDPG

سریع بودن PDDPG 1.7 برابر DPG

کمتر بودن سه درصدی هزینه PDDPG نسبت به DPG در زمان همگرایی

آنالیز محاسباتی روش ها

TABLE II

COMPUTATIONAL PERFORMANCE OF THE EXAMINED DRL METHODS

Method	DQN	DPG	PDDPG
Total training time (s)	1,008	1170	949
Number of episodes	8,000	13,000	6,200
Average training time per episode (s)	0.13	0.08	0.15

سریع بودن PDDPG از به علت
مکانیزم تجربه اولویت دار
سرعت پایین DPG به علت وجود
واریانس بالا در تخمین گرادیان

زمان آموزش هر اپیزود
آموزش یک DNN توسط DPG
نیاز به آموزش یک DNN در هر گام توسط
DQN

جمع بندی

- ارائه مساله مدیریت انرژی برای سیستم های چندحاملی خانگی
- در نظر گرفتن تصادفی و متغیر بودن تقاضای انرژی، قیمت ها و تولید PV
- ترکیب روش PDDPG با مکانیزم تکرار تجربه دوباره دارای اولویت
- 19 درصد کم تر بودن هزینه روزانه PDDPG نسبت به DQN و 8 درصد نسبت به DPG

پیشنهاد کار آینده

- بررسی سیستم های چند حاملی با ابعاد بزرگ تر: منابع انرژی (برق، آب، گاز، هیدروژن، گرما)، تجهیزات مبدل انرژی (پمپ های آب، چیلرهای برقی)
- در نظر گرفتن مدیریت انرژی برای چندین سیستم چندحاملی
- تعمیم PDDPG برای نمایش مقاوم بودن آن برای تغییرات زیاد شرایط زیست محیطی همانند شرایط بد آب و هوا، کار نکردن صفحه خورشیدی و ذخیره ساز انرژی و افزایش یا کاهش تعداد ساکنان ساختمان ها