IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 16, 2021, accepted May 26, 2021, date of publication June 8, 2021, date of current version June 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3087491

Resilience Microgrid as Power System Integrity Protection Scheme Element With Reinforcement Learning Based Management

LILIA TIGHTIZ, (Student Member, IEEE), AND HYOSIK YANG, (Member, IEEE)
Department of Computer Engineering, Sejong University, Seoul 05006, South Korea

Corresponding author: Hyosik Yang (hsyang@sejong.edu)

This work was supported by the Technology Development Program to Solve Climate Changes through the National Research Foundation of Korea (NRF) by the Ministry of Science and ICT under Grant NRF-2021M1A2A2065447, and in part by Korea Electric Power Corporation under Grant R17XA05-2.

ABSTRACT The microgrid is a solution for integrating renewable energy resources into the power system. However, overcoming the randomness of these nature-based resources requires a robust control system. Moreover, electricity market participation and ancillary service provision for the utility grid are other aspects, although intensify microgrid penetration makes its environment interactions more complex. Reinforcement learning is a technique vastly applied to such an intricate environment. Hence, in this paper, we deployed deep deterministic policy gradient and soft-actor critic methods to solve the high-dimensional, continuous, and stochastic problem of the microgrid's energy management system and compared the performance of two methods. Additionally, we developed the microgrid interactions with the utility grid as a participant of system integrity protection schema responding promptly to the utility grid protection requirements based on its reliable available resources. Moreover, we applied actual data of Gasa Island microgrid in Korea to prove the efficiency of proposed method.

INDEX TERMS Energy management system, deep deterministic policy gradient, soft actor-critic, system integrity protection schema.

ABBREVIATIONS

BEMS	Building Energy Management Systems
BESS	Battery Energy Storage Systems
DDPG	Deep Deterministic Policy Gradient
DG	Diesel Generator
DNN	Deep Neural Networks
DPG	Deterministic Policy Gradient
DDQN	Double Deep Q-Network
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
EMS	Energy Management Systems
ESS	Energy Storage Systems
EV	Electric Vehicles
FACTS	Flexible AC Transmission Systems
IED	Intelligent Electronic Devices
KPX	Korean Power Exchange Company
LMP	Local Marginal Price
MDP	Markov Decision Process
PCC	Point of Common Coupling
PV	PhotoVoltaic cells
SIPS	Power System Integrity Protection Schema
RER	Renewable Enegry Resources
RL	Reinforcement Learning
SAC	Soft Actor-Critic
SoC	State of Charge
SPG	Stochastic Policy Gradient
TRPO	Trust Region Policy Optimization
WT	Wind Turbine

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

NOMENCLATURE

α	Initial temperature coefficient in SAC algorithm.
γ	Bellman equation's discount facotr.
ε	The probabilty of choosing to explore.
η_{ch}	BESS charging loss (%).

η_l	BESS leakage loss (%/month).
ρ	BESS lifetime degradation coefficient.
τ	Target smoothing coefficient in SAC algorithm.
a_{1i}, a_{2i}, a_{3i}	Coefficients of each DG's fuel cost.
$\mathcal{B}$	SIPS coefficient in reward function.
$Cost_{invest}$	Initial investment for BESS provision (₩).
C_i	Each element of microgrid's energy production cost.
E_i	Energy provided by each element of microgrid.
E_B	Amount of stored energy in BESS.
$E_{B,rated}$	Nominal amount of stored energy in BESS.
K	Number of DG units in the microgrid domain.
k_i	Reliability coefficient of each energy provider in SIPS.
LMP	Hourly local marginal price (₩/kWh).
M	Number of WT units in the microgrid domain.
N	Number of PV units in the microgrid domain.
$N_{ch,dch}$	BESS's number of full-charge and discharge cycle number.
$P_{B,ch,max}$	Maximum of BESS charging rate (kW).
$P_{B,dch,max}$	Maximum of BESS discharging rate (kW).
P_{Demand}	The amount of loads need to supply (kW).
P_{DG}	DG output power (kW).
$P_{DG,rup}$	Ramp-up rate of DG (kW).
$P_{DG,rdown}$	Ramp-down rate of DG (kW).
P_{net}	The amount of surplus and shortage of power (kW) in the microgrid without BESS and DG.
$P_{Shedding}$	The amount of shedding loads (kW).
$P_{Purchase}$	The amount of purchased power from the utility grid (kW).
P_{PV}	PV output power (kW).
P_{Sell}	The amount of sold power to the utility grid (kW).
P_{WT}	WT output power (kW).
R	Number of BESS units in the microgrid domain.
SoC	BESS state of the charge.
Δt_i	Time duration of each microgrid element contribution in energy production.
$T_{DG,down}$	DG's minimum down time.
$T_{DG,up}$	DG's minimum up time.
$T_{DG,on}$	Number of hours that DG is on.
$T_{DG,off}$	Number of hours that DG is off.

I. INTRODUCTION

Inexhaustible, nature-friendly, and high-efficiency specifications of renewable energy resources (RER) will turn them into a dominant source of power generation over the globe shortly. However, the power generation of RER inherited stochastic characteristics of its nature-origin [1]. Utilizing RER with assists of energy storage systems (ESS) and conventional generators lead to control uncertainty of RER. Microgrid, which is an independent subset of the power grid, employs ESS and conventional generators to accelerate the penetration of RER. Given the autonomous feature of microgrid, it is primarily widely applied in remote places. With the advent of smart inverters, enhanced control systems, and electricity markets, the microgrid integrates into the utility grid [2]. This schema calls for an intelligent control system called the energy management system (EMS) that inspects the microgrid's technical priorities and stakeholders' objectives and finds optimal future action of the system [3]. The research area of EMS in the microgrid has risen application of a wide variety of optimization methods. Deterministic and heuristic optimization techniques, classical machine learning techniques, and their combinations, such as fuzzy logic expert [4], dynamic programming [5], non-linear programming [6], and meta-heuristic methods [7], [8] have been applied to improve the EMS performance. Uncertainty in load profile, RER power generation, market price, and state of the charge (SoC) of ESS make EMS planning a non-deterministic and continuous problem. On the other hand, it is a high-dimension issue because of the number of actors and their behavior. The aforementioned methods will not satisfy the non-deterministic and high-dimension characteristics of EMS [9]. To combat this complexity trend has been arisen to utilize reinforcement learning (RL) techniques in the EMS arrangement. RL is a framework in which an agent for each action receives a reward from the surrounding environment, and based on that reward, learns to do optimal actions. Since Q-learning is a model-free method, it is one of the fascinating RL methods for EMS planning to fulfill the microgrid explicit model inaccessibility. This model learns from real-time data while future rewards or state situations of the system are unknown. Q-learning applied in building energy management systems (BEMS) representing the microgrid [10], [11]. Battery scheduling is provided based on uncertainty in wind turbine (WT) power generation as a power source with the principal purpose of less electricity purchase from the utility grid in [10]. The authors in this paper declined uncertainty on load and electricity market price and considered scheduling for 2 hours a day ahead. It is worth nothing that in this study the microgrid does not export energy to the main grid. Conversely, Kim and Lim [11] scheduled a real-time EMS of smart building storages as a prosumer while the main objective is reducing energy cost.

Though, to some extent, Q-learning achieved success in planning EMS, exploiting microgrid elements capabilities and considering their uncertainty is a high-dimensional issue which is where Q-learning traps in the curse of dimensionality. Deep Q-Network (DQN) [18] is a combination of deep neural networks (DNN) and RL that offers scalability to the microgrid's EMS arrangement. Since it utilizes neural networks for approximating the value of states instead of Q-learning tabular representation of state value supports high-dimensional problems. References [12] and [13] in the same approach applied convolutional neural network in DQN for estimating the value of states to schedule battery performance and considered the microgrid is an independent grid that does not have any energy transactions with the utility grid. In another effort [14], the uncertainty of load profile, RER, and market price were considered in providing cost-effective EMS in the microgrid by applying DQN. The authors in this paper assessed their method efficiency deploying real data from California independent system operators. Albeit DQN outperforms Q-learning in a stochastic environment, it still suffers from instability in the training network because of the correlation between the target value and estimated value. Double DQN (DDQN) is a solution to this problem which offers two separated neural networks for action selection and action evaluation [19]. To combat the overestimation drawback of DQN, in [15], DDQN was applied to optimize the performance of the microgrid battery EMS community. Operation of ESS scheduled through learning process whereby in the grid-connected mode of microgrid, the cost of power generation minimized and in islanded mode as much as critical load supplied. The author in this paper regarded the electricity market prices, loads, and RERs as uncertain elements of the environment states and extended their presented method to the multi-micro grid system.

Q-learning and its enhanced derivate methods, including DQN and DDQN are value-based learning methods. They have appropriate performance for discrete actions, and the slow rate of policy change makes them dedicated to the drawback of overestimation. Deep learning is also applied to the other class of RL, which is the policy-based method and formed policy gradient-based techniques. Policy gradient-based method classified into stochastic and deterministic approaches. Stochastic policy gradient (SPG) surplus value-based method in convergence speed and solving high dimension action environment problems even though there is a risk of convergence to the local maximum. Actor-critic is a trade-off between value-based and policy-based procedures in which the actor defines action based on policy and critic evaluates the value of each action. There is a wide range of SPG and deterministic policy gradient (DPG) based actor-critic techniques. Deep deterministic policy gradient (DDPG) is an example of DPG-based methods, while soft actor-critic (SAC) is an SPG-based technique. Since the microgrid has an uncertain environment, SGP methods can fit EMS requirements. Mocanu *et al.* [16] analyzed the superior of DPG to DQN in BEMS whereas minimizes the cost of energy and peak demand. BEMS in this system considers the uncertainty of electric vehicles (EV) owner while the role of ESS as a fundamental member of EMS ignored. Nakabi and Toivanen [17] considered the role of demand response in the microgrid EMS and categorized loads into price-based and temperature-based participating in demand response through peak-time shifting and arbitrage, respectively. Authors in this paper, after providing a Markov model of the microgrid provide a comprehensive comparison of different deep reinforcement learning (DRL) methods in solving the microgrid EMS problem, including SARSA, DDQN, Actor-Critic, PPO, and A3C and revealed when PPO and A3C follow a semi-deterministic approach through action selection from replay experience buffer in exploration, they can enhance learning system convergence. Table 1 summarizes a comparison between the present paper and related works.

Driven by maximizing profit for the microgrid's stakeholders, the abovementioned studies conducted their method with the hypothesis of bi-directional energy trading of the microgrid in the point of common coupling (PCC). In this paper, as well as following the previous research objective of minimizing the cost of energy production of the microgrid, we granted the EMS scenario, with the novel approach, to acts as a member of the power system integrity protection schema (SIPS). Enormous financial loss and negative social effects of interruption in power system services encouraged integrity in power system protection. SIPS is a collection of measurements, monitoring, communications, decisions, and actions to protect the stability of the power system in contingencies [20]. It is estimated that the future power grid will be a multi-microgrid system due to the ever-increasing RERs utilization, in particular, in the form of the microgrid. This fact reveals the necessity of the microgrid examination as a member of SIPS. Therefore, our DRL-based algorithm for EMS determines the microgrid to work in islanding, consumer mode, or generation mode to maximize the association of reliable elements in response to the grid SIPS signal. As shown in Table 1, DDPG and SAC can defeat the vulnerability of the other methods in not supporting high-dimensional and continuous specifications of the microgrid environment. Especially when the microgrid is granted to the SIPS member of the power system, the complexity degree of the microgrid environment will increase. Due to the superior of DDPG and SAC in supporting microgrid environment specifications, we will deploy both methods to solve the EMS problem and compare the results. In a nutshell, the main contributions of this paper are as follows:

- Determine the microgrid structure, elements, and constraints to arrange Markov decision process (MDP) of the microgrid considering novel control system for associating in SIPS.
- Propose DRL framework for EMS of the microgrid based on both DDPG and SAC methods to compare their performances.
- Investigate the accuracy of our technique in normal and SIPS situations with different scenarios.

The remainder of this paper will be as follows: Section II provides the microgrid structure and constraint functions. This section also reveals the concept and application requirements of SIPS and the microgrid's role in that. We devote Section III to the MDP arrangement of the microgrid and solution algorithm. We present different scenarios to estimate our method performance and analyze results in Section IV. Ultimately, this paper ceases in Section V as a conclusion.

TABLE 1. Comparison of related works with the present work.

Ref	Main objective	Microgrid elements	Microgrid type						Applied method			
			Islanded	Prosumer					Main method	Benchmark method	Evaluation with real data set	Comments on main RL applied method
				Electricity market participation arrangement								
				Import power from the grid	Export power to the grid		Market price					
					Market Participant	Ancillary Service Provider	Constant	Stochastic				
[10]	Optimal battery scheduling to reduce dependency to the utility grid	WT, BESS	✗	✓	✗	✗	✓	✗	Q-Learning	✗	✗	Not support complex, stochastic, and continuous environment of microgrid because of curse of dimentionality problem
[11]	Building EMS to minimize energy cost	PV, BESS, Super Capacitor, Responsive loads (EVs)	✗	✓	✓	✗	✗	✓	Q-Learning	✗	✓	Not support complex, stochastic, and continuous environment of microgrid because of curse of dimentionality problem
[12]	Islanded microgrid energy management system	PV, BESS, Hydrogen storage device	✓	✗	✗	✗	✗	✗	DQN	✗	✓	Not support stochastic, and continuous environment of microgrid because of overestimation problem
[13]	Islanded microgrid energy management system	DG, PV, BESS, Hydrogen storage device	✓	✗	✗	✗	✗	✗	DQN	Mixed integer programing	✓	Not support stochastic and continuous environment of microgrid because of overestimation problem
[14]	Microgrid EMS concerning load, RES, and market price	DG, WT, PV, BESS, Fuel cells	✗	✓	✓	✗	✗	✓	DQN	Q-Learning, Fitted Q-Learning	✓	Not support stochastic and continuous environment of microgrid because of overestimation problem
[15]	Community of BESS scheduling in microgrid	DG, WT, PV, BESS	✗	✓	✓	✗	✗	✓	DDQN	✗	✗	Not suited microgrid high-dimentional environment
[16]	Building energy management system scheduling to minimizes the cost of energy and peak demand	RES, BESS, Responsive loads (Home appliances, EVs)	✗	✓	✓	✗	✗	✓	DPG	DQN	✓	DPG: Trap in local maximum rather than finding global optimum
[17]	Role of demand response in the microgrid EMS	WT, Community of BESS, price-based and temperature-based loads	✗	✓	✓	✗	✗	✓	DQN	SARSA, DDQN, REINFORCE, PPO, PPO++, A3C, actor-critic	✗	Complexity of hyper parameters selection for actor-critic based methods
Present paper	Scheduling EMS for Microgrid to maximum profit in grid-connected mode and maximum contribution in SIPS	DG, WT, PV, BESS, Responsive loads	✗	✓	✓	SIPS Contribution	✗	✓	DDPG	SAC	✓	DDPG-Pros: support continuous and high-dimentional microgrid environment DDPG-Cons: Not support stochastic policy hence there is one action for each state SAC-Pros: More effecient action space due to supporting stochastic policy and stable training process SAC-Cons: Time-consuming training process

II. MICROGRID STRUCTURE

In this paper, we considered the microgrid has the capability of energy transaction with the utility grid at the PCC. Fig.1 shows that the understudy microgrid includes photovoltaic cells (PV), WT, diesel generator (DG), ESS, and critical and non-critical loads. Fig.1 also delineates the microgrid control levels and microgrid interactions with the supervisory sector of the power system to receive commands and data, such as electricity market prices and SIPS alarms.

Bidirectional flow of both power and information in the power system due to the smart grid implementation along with the introduction of digital technologies to power system protection industry enhanced power system protection from applying electromagnetic relays to intelligent electronic devices (IED) utilization. Apart from facilitating the integration of new equipment to the power system, such as RER and flexible AC transmission systems (FACTS), aforementioned developments provide the possibility of planning schemes for the protection of the power system in narrow boundaries, which are known as SIPS. Defending power system stability and integrity, preserving critical equipment against damage, and preventing blackout extension in the course of disturbances are notable SIPS missions. SIPS implemented by utilizing measurement and control devices and decision-maker algorithms in different levels of the power system.

SIPS offers several scenarios dealing with abnormal situations according to system historical performance information and its present condition. Although the principal role of SIPS is restricting blackout extension, under-voltage, under-frequency, thermal overloading, congestion, and transient instability are other sample issues on that SIPS can assist power system. SIPS responds to these issues by applying techniques, such as load shedding, system topology reconfiguration, generator rejection, and so forth. SIPS applied locally to protect power system distribution and transmission networks. This local protection usually has a flat architecture since decision-maker elements and measurements are located in distribution or transmission substations.

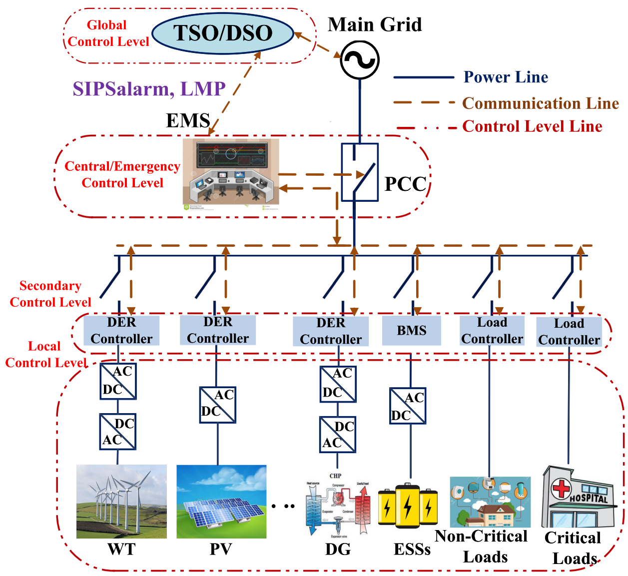


FIGURE 1. Microgrid elements and control leves.

Furthermore, SIPS with extended horizon practiced to large utility or interconnected utility protection. This extended schema has different styles, such as hierarchical, centralized, or distributed [21]. In our scenario, regardless of SIPS architecture, we examined the microgrid as membership of SIPS and power system assistant during contingencies. Microgrid's EMS unit receives SIPS signals and replies to them based on the availability and reliability of resources specified in Section III.

Each element of the microgrid has its model and constraints, which should be regarded in planning EMS, which we discussed in detail as follows.

A. RERs CHARACTERISTICS

WT and PV are common RERs applied in the microgrid. Heuristic-based optimization methods were popular methods in modeling WT and PV to predict their power output. Given deploying the model-free technique in this study, we examined stochastic characteristics of RERs and evoked their output power to train our model based on historical data. Each RERs have a minimum and maximum output power limitation.

$$P^i_{PV,min} \leq P^i_{PV}(t) \leq P^i_{PV,max}, \quad 1 \leq i \leq N \tag{1}$$
$$P^i_{WT,min} \leq P^i_{WT}(t) \leq P^i_{WT,max}, \quad 1 \leq i \leq M \tag{2}$$

where N and M represent the number of PVs and WTs in the microgrid and P_{PV} and P_{WT} are their output power, respectively.

RERs are the principal suppliers of customers in the microgrid domain. Hence, we decline the cost of energy delivered by RERs. Surplus generation of RERs will be reserved in ESS to meet power demand in their unavailability, contingency requirements, and electricity market.

B. DG MODEL

DG is the other source of power generation in the microgrid serving demand in RERs absence. We assume the microgrid has K unit of DG; each unit generates P_{DG} at each time t. The energy generation cost of DG is described according to (3), similar to the traditional fossil-fuel power plants cost function.

$$Cost_{DG^i}(t) = a_{1i} + a_{2i}P^i_{DG}(t) + a_{3i}(P^i_{DG}(t))^2, \quad 1 \leq i \leq K \tag{3}$$

where a_1, a_2, and a_3 are coefficients of generator fuel cost. Additionally, (4) shows the generation restriction of DG while (5) and (6) represent ramp-down and ramp-up constraints of DG, respectively [22].

$$P^i_{DG,min} \leq P^i_{DG}(t) \leq P^i_{DG,max} \tag{4}$$
$$P^i_{DG}(t-1) - P^i_{DG}(t) \geq P_{DG^i,rdown} \; \text{if } P^i_{DG}(t-1) > P^i_{DG}(t) \tag{5}$$
$$P^i_{DG}(t-1) - P^i_{DG}(t) \geq P_{DG^i,rup} \; \text{if } P^i_{DG}(t) > P^i_{DG}(t-1) \tag{6}$$

The DG state can change in a certain time step with respect to its minimum up and down time characteristics as follows.

$$(T_{DG^i,on}(t-1) - T_{DG^i,up})(S_{DG^i,t-1} - S_{DG^i,t}) \geq 0 \tag{7}$$
$$(T_{DG^i,off}(t-1) - T_{DG^i,down})(S_{DG^i,t} - S_{DG^i,t-1}) \geq 0 \tag{8}$$

where $T_{DG,up}$ and $T_{DG,down}$ determine minimum up and down time, respectively. S_{DG} is a binary value that shows the DG is on or off. Given the DG's reliable performances in contingencies, such as frequency control, these limitations can be declined for units participating in SIPS [23].

C. ESS CONSTRAINTS

ESS in the under-study microgrid is BESS with R numbers not only provide the backup power for the microgrid also assists the system stability. If we consider coefficient η_l describing leakage loss and η_{ch} as charging loss, BESS's dynamic model in the time interval Δt is described as follows.

$$E^i_B(t) = E^i_B(t-1) - [P^i_B(t) + \eta_{ch}P^i_B(t) + (\eta_l|P^i_B(t)|)]\Delta t, \quad 0 \leq i \leq 1 \tag{9}$$
$$SoC(t) = E_B(t)/E_{B,rated} \tag{10}$$

BESS component delivers power with constraints according to (11), (12), and (13). Where SoC_{min} and SoC_{max} are the minimum and maximum of battery SoC. Additionally, $P_{B,ch,min}$ and $P_{B,dch,max}$ are minimum and maximum charging and discharging rate of the BESS, respectively.

$$-P^i_{B,ch,max}(t) \leq P^i_B(t) \leq P^i_{B,dch,max} \tag{11}$$
$$P^i_{B,ch,max}(t),\, P^i_{B,dch,max}(t) \geq 0 \tag{12}$$
$$SoC^i_{min} \leq SoC^i(t) \leq SoC^i_{max}, \quad 1 \leq i \leq R \tag{13}$$

The cost of power delivered by BESS is concerned with (14).

$$Cost_B^i = [P_B^i(t) + E_B^i(t)\eta_l]\rho\Delta t \tag{14}$$

$$\rho = Cost_{invest}/(E_{B,rated}.N_{ch,dch}) \tag{15}$$

where ρ is the battery lifetime degradation factor, $E_B(t)$ is the amount of stored energy in BESS at time t, $E_{B,rated}$ is the nominal amount of stored energy in BESS, the initial investment for BESS provision is $Cost_{invest}$ and $N_{ch,dch}$ is the full charge and discharge cycle number [22].

D. POWER AND LOAD CONSTRAINTS

The microgrid in our scenario supplies critical and non-critical loads in which demand response scheduling is implemented on non-critical loads. Furthermore, local marginal price (LMP) is a mechanism that will be applied to conduct economic calculation of bidirectional microgrid energy transactions in the PCC. We assume the local electricity market in one hour ahead offers LMP. Scheduling the microgrid involvement in the electricity market should meet electricity generation and consumption balance as follows.

$$P_{net}(t) = P_{Demand}(t) - (\sum_{i=1}^{N} P_{PV}^i(t) + \sum_{i=1}^{M} P_{WT}^i(t)) \tag{16}$$

$$P_{net}(t) - (\sum_{i=1}^{K} P_{DG}^i(t) + \sum_{i=1}^{R} P_B^i(t) + P_{Sell}(t) - P_{Purchase}(t) + P_{Shedding}(t)) = 0 \tag{17}$$

$$P_{Sell}(t).P_{Purchase}(t) = 0 \tag{18}$$

where $P_{Demand}(t)$, $P_{Sell}(t)$, $P_{Purchase}(t)$, and $P_{Shedding}(t)$ are the microgrid's consumer power demand, the amount of selling and purchasing power to and from the utility grid, and the amount of sheddable load at each time t, respectively. Equation (18) fulfills avoiding the microgrid's selling and purchasing energy at the same time.

III. APPLIED METHOD

A. MICROGRID MARKOV DECISION PROCESS

In RL methods, an agent tries to learn the best policy concerning each action's reward receives from the environment. The EMS is an agent interacts with the environment. This environment includes the microgrid's elements and the utility grid supervisory system. EMS as an agent tries to learn how to minimize the cost of energy production in a normal situation and maximize contribution in SIPS situations. We modeled this environment by the MDP. Therefore, the first step of RL process implementation is the MDP arrangement for the microgrid. MDP is defined by the state, action, transition function, and reward in the form of 4-tuple $\{\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}\}$, which are state space, action space, transition function, and reward function, respectively, and discussed as follows.

1) STATES

The states clarify the environment's elements status for the agent. Since our priority is deploying RERs, their status evoked through P_{net} calculated by (16) is one of the parameters that show our environment's state. The other states are SoC of the battery, LMP is issued every hour by the market operator, the amount of load participating in load shedding, and time. Additionally, the SIPS situation is the other state reminds the microgrid duties as membership of power system protection planning. Hence, the state space $\mathcal{S}$ determined as follows.

$$\mathcal{S} = \{P_{net}, SoC, LMP, SIPS_{sign}, P_{shedding}, time\} \tag{19}$$

2) ACTIONS STATE

Since EMS in each state should respond to load, electricity market, and SIPS alarm, the action space is a set of activities able to satisfy these requirements. As discussed before, equation (17) guarantees in each state, the microgrid will meet demands inside the microgrid by utilizing RERs, BESSs, DGs, and load shedding. In response to the market price, EMS determines the microgrid should purchase power from the utility grid or sell surplus power to that. The microgrid in the SIPS situation, by taking actions, including islanding, BESS discharging, load shedding, and DG utilization reacts to the protection requirements of the utility grid. Given the primary mission of the microgrid is satisfying loads, other actions are subsets of $\mathcal{A}_{load}$. For this reason, we gather all actions in a unit action space according to (21).

$$\mathcal{A} = \{A_{load}, A_{Market}, A_{SIPS_{sign}}\} \tag{20}$$

$$\mathcal{A} = \{DG_{action}, Charging_{BESS}, Disharging_{BESS}, Islanding, Purchasing, Selling, loadShedding\} \tag{21}$$

3) TRANSITION FUNCTION

Following stochastic and partially observed microgrid environment characteristics, the current state $\mathcal{S}_t$ and current action a_t are the only parameters that determine the next step. Considering the uncertainty of RERs output power, demand, LMP, and occurring SIPS situations that determine our environment state, the transition function probability, defined in (22), is unknown in each time step. This fact outstands the priority of utilizing a model-free RL algorithm in solving the EMS problem of the microgrid.

$$\mathbb{P}(\mathcal{S}_{t+\Delta t}, \mathcal{S}_t, a_t) = \mathbb{P}[\mathcal{S}_{t+\Delta t}|\mathcal{S}_t, a_t] \tag{22}$$

4) REWARD FUNCTION

In RL, the environment by granting reward to each agent's action leads the system to discover the best actions to achieve its objectives. In our scenario, the microgrid participating in the electricity market and SIPS follows two separated goals shown as bellows.

$$cost = \begin{cases} Min[\sum_{i=1}^{n} \Delta t_i * C_i * E_i], & \text{if } Normal\ Situation \\ Max[\sum_{i=1}^{n} \Delta t_i * E_i * k_i], & \text{if } SIPS\ Alarm \end{cases} \tag{23}$$

where;

n: number of participants in the microgrid's EMS,

$E_i \in \{E_{RER}, E_{DG}, E_{Sh}, E_{BESS}\}$,
$C_i \in \{C_{RER}, C_{DG}, C_{Sh}, C_{BESS}\}$,
$\Delta t_i \in \{\Delta t_{RER}, \Delta t_{DG}, \Delta t_{Sh}, \Delta t_{BESS}\}$,
$k_i \in \{0, l, 1\}, 0 \prec l \prec 1$,
k: Reliabilty coefficient of each participant in SIPS,
E_{RER}: Energy delivered by RERs,
E_{DG}: Energy delivered by DGs,
E_{Sh}: Energy conserved by sheddable loads,
E_{ESS}: Energy delivered by BESSs,
C_{RER}: Price of generated energy by RERs,
C_{DG}: Price of generated energy by DGs,
C_{Sh}: Price of sheddable energy,
C_{BESS}: Price of stored energy,
Δt_{RER}: Time duration of access to RERs,
Δt_{DG}: Time duration of access to DG,
Δt_{Sh}: Time duration of access to sheddable loads,
Δt_{BESS}: Time duration of access to BESS,

Given the system objectives, the reward in the normal situation is the revenue obtained from market participation after subtracting the cost of the power supplier's performance. Because the microgrid maximizes cooperation with the utility grid, in the SIPS situation, we withdraw from acquiring revenue by devoting zero to coefficient $\mathcal{B}$ to consider only the reliability of the system suppliers. To generalize applying reward to the microgrid actions, we take into account that all members of action space in (21) will lead to the system work in different situations according to (26), (27). By devoting weight to each system action, the reward function is as below.

$$Revenue = \mathcal{B} * [\sum_{i=1}^{n} E_i * LMP_i], \tag{24}$$

where,

$$\mathcal{B} = \begin{cases} 0, & \text{if } SIPS\ Alarm; \\ 1, & \text{if } Normal\ Situation; \end{cases} \tag{25}$$

for Normal Situations:

$$\mathcal{A} = \begin{cases} 0, & \text{Islanded;} \\ 1, & \text{Grid-Connected(Producer);} \\ -1, & \text{Grid-Connected(Consumer);} \end{cases} \tag{26}$$

for SIPS Situations:

$$\mathcal{A} = \begin{cases} 1, & \text{Commissioning;} \\ -1, & \text{Not-Commissioning;} \end{cases} \tag{27}$$

$$Reward = \mathcal{A} * |Revenue - Cost|, \tag{28}$$

B. SOLUTION ALGORITHM

As discussed before, there are some difficulties in applying value-based RL methods, such as weak convergence performance for high-dimensional problems. However, the recent enhances in these methods, such as DQN, tried to conquer this issue. In DQN, instead of creating a table to estimate value-function, employ a DNN for this estimation and provides i.i.d input for DNN by applying a buffer of estimated Q-values called replay memory. This method also deploys two separated DNN for Q-value estimation working online and target network updated by the online network after a predefined time step. Although these signs of progress make value-based have an appropriate performance for discrete actions, the slow rate of policy changing makes them dedicated to the drawback of overestimation. DDPG is an actor-critic and deep learning-based algorithm. DDPG takes advantage of the aforementioned solutions, namely, replay memory and deploying four separated DNN, including Q-network (θ^Q), deterministic policy function (θ^μ), target Q-network ($\theta^{Q'}$), and target policy network ($\theta^{\mu'}$) [24]. In each trajectory, Bellman equation (29) updates the value. Since DDPG follows the actor-critic method, the next state Q-values of target value networks, i.e., actor and critic, are calculated according to (30) and (31). Consequently, the actor policy is updated by minimizing the loss function according to equation (32). Moreover, DDPG uses batch normalization to solve internal covariate shift difficulty. DDGP algorithm is according to 1.

$$y_i = r_i + \gamma Q'(\mathcal{S}_{i+1}, \mu'(\mathcal{S}_{i+1}|\theta^{\mu'})|\theta^{Q'}) \tag{29}$$

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}, \quad \tau \ll 1 \tag{30}$$

TABLE 2. Case study microgrid specifications.

DG	$P_{min}(kW)$		75
	$P_{max}(kW)$		330
	$P_{DG,ramp-down}(kW/step)$		120
	$P_{DG,ramp-up}(kW/step)$		120
	$T_{DG,up}(h)$		2
	$T_{DG,down}(h)$		2
	Quadratic Coefficients	a_1	1.3
		a_2 (₩/kW)	0.0304
		a_3 (₩/kW^2)	0.00104
BESS	SoC_{min}		20%
	SoC_{max}		95%
	$P_{ch,max}(kW)$		70
	$P_{dch,max}(kW)$		70
	η_l		3%/month
	ρ(₩/MW)		100
DRL Hyperparameters	Learning rate		0.0001
	Discount factor (γ)		0.9
	Replay buffer size		50,000
	Number of training episodes		5,000
	Mini-batch size		128
	Number of hidden layer		2
	Initial temperature (α)		0.05
	λ_Q, λ_π, λ_α		0.0003
	Target smoothing coefficient (τ)		0.005
	Activation Function		ReLU
	Optimizer		SGD

$$\theta^{\mu} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'}, \quad \tau \ll 1 \tag{31}$$

$$\frac{1}{N}\sum_{i}(y_i - Q(\mathcal{S}_i, a_i|\theta^Q))^2 \tag{32}$$

Algorithm 1: DDPG Algorithm

Initialization randomly critic network Q($\mathcal{S}$, $a|\theta^Q$) and actor $\mu(\mathcal{S}|\theta^\mu)$ with weights θ^Q and θ^μ ;
Initialize target network Q' and μ' with weights $\theta^{Q'} \leftarrow \theta^Q, \theta^{\mu'} \leftarrow \theta^\mu$;
Initialize replay buffer R;
for *batchsize* $\in \{1, \ldots, M\}$ **do**
  Initialize a random process $\mathcal{N} \in$ action exploration noise ;
  Receive initial observation state $\mathcal{S}_1$;
  for $t \in \{1, .., T\}$ **do**
    Select action $a_t = \mu(\mathcal{S}|\theta^\mu) + \mathcal{N}_t$, according to the current policy and exploration noise Execute action and observe reward r_t and observe new state $\mathcal{S}(t+1)$;
    Store transition ($\mathcal{S}_t, a_t, r_t, \mathcal{S}_{t+1} \in$ R;
    Sample a random minibatch of transitions ($\mathcal{S}_t$, $a_t, r_t, \mathcal{S}_{t+1}$ from R;
    Set according to Update critic by minimization the loss according to (32);
    Update the actor policy using the sampled policy gradient by:

$$\nabla_{\theta^\mu}\mathcal{J}(\theta) \approx \frac{1}{N}\sum_{i}[\nabla_a Q(s, a|\theta^Q, s = s_i, a = \mu(\mathcal{S}_i)) \nabla_{\theta^\mu}\mu(\mathcal{S}|\theta^\mu, s = s_i)] \tag{33}$$

    Update the target networks according to (30), (31) ;

Deterministic characteristic of the actor which conducts interaction with Q-function makes DDPG dedicated to instability. SAC is an actor-critic RL algorithm introduced to solve DDPG drawbacks, such as being hyperparameter and unstable, by adopting a stochastic actor. In addition to deploying stochastic policy to obtain stability characteristics of on-policy approaches, such as trust region policy optimization (TRPO) and PPO, SAC employs the replay buffer of DDGP as a deterministic procedure to acquire efficient samples by reviewing previous operation instances [25]. The backbone of SAC is entropy regularization to explore random actions, which is interpreted as a soft and large choice of actions and avoid converging to local minima [26]. The stochastic policy and entropy are shown in (34) and (35).

$$\pi^* = argmax_{\pi}\sum_{t}(\mathbb{E}_{(s_t,a_t)} \sim_{\rho\pi} \times [r(s_t, a_t) + \alpha\mathcal{H}(\pi(.|s_t))] \tag{34}$$

$$H(\pi(.|s_t)) = \mathbb{E}_{a\sim\pi(.|s)}[-log(\pi(a|s))] \tag{35}$$

Algorithm 2: SAC Algorithm

Initialization policy network with weight ϕ and parameter λ_π ;
Initialize target network Q with weights $\overline{\theta}_1 \leftarrow \theta_1$, $\overline{\theta}_2 \leftarrow \theta_2$ and parameterλ_Q;
Initialize replay buffer $\mathcal{D}$;
for $k \in \{1, 2, \ldots\}$ **do**
  for $t \in \{1, 2, \ldots\}$ **do**
    Sample $a_t \sim \pi_\theta(a_t|\mathcal{S}_t)$;
    $\mathcal{S}_{t+1} \sim \rho_\pi(\mathcal{S}_{t+1}|\mathcal{S}_t, a_t)$;
    $\mathcal{D} \leftarrow \mathcal{D} \cup \{\mathcal{S}_t, a_t, r(\mathcal{S}_t, a_t)\}, \mathcal{S}_{t+1}$;
    for *each update of gradient* **do**
      $\theta_i \sim \theta_i - \lambda_Q \nabla_{\theta_i} \mathcal{J}_Q(\theta_i) fori \in \{1, 2\}$
      $\phi \sim \phi - \lambda_\pi \nabla_\phi \mathcal{J}_\pi(\phi)$
      $\alpha \sim \alpha - \lambda_a \nabla \mathcal{J}(\alpha)$
      $\overline{\theta}_i \sim \tau\overline{\theta}_i + (1-\tau)\overline{\theta}_i fori \in \{1, 2\}$

SAC upgrades MDP by input entropy as a fifth element of MDP tuple set. This approach added entropy maximization to the traditional Bellman equation objective, previously defined as maximizing the expected cumulative reward. Hence, the updated Bellman equation based on the stochastic policy is as follows.

$$V(s_t) = \mathbb{E}_{a\sim\pi_\phi(.|s_t)}[Q_\theta(s_t, a) - \alpha log(\pi_\phi(a|s_t))] \tag{36}$$

where α is the temperature parameter. In SAC, soft critic evaluates soft Q-function of policy and actor by using critic's evaluation improves maximum entropy policy and temperature coefficient adjusts entropy for maximizing entropy policy. SAC deploys conditional Gaussian density to create trackable stochastic policy, therefore the action according to (37) updated by mean and variance of samples where initialize by $\epsilon \sim \mathcal{N} \in (0, 1)$.

$$a_\theta(s, \epsilon) = tanh(\mu_\theta(s) + \delta_\theta(s).\epsilon) \tag{37}$$

Algorithm 2 represents SAC. Fig. 2 represents an overview of the microgrid environment and applied solution algorithms. Continuous and model-free microgrid structure motivates applying both DDPG and SAC methods in scheduling EMS in the microgrid. Hence, in this paper, we employ both methods and compare their results.

IV. CASE STUDY

For our proposal examination, we deployed the Gasa Island microgrid specifications and scheduled EMS for one hour ahead based on our method. Gasa Island is an RER-based standalone microgrid located in the southeast of the Korean peninsula. This island, according to Table 2, equipped with 314kW PV, 400kW WT, 3MWh BESS, and 300 kW DG to supply mainly agricultural and residential loads with 173kW peak load. There is an EMS unit to control and monitor Gasa Island microgrid elements. We deployed Gasa

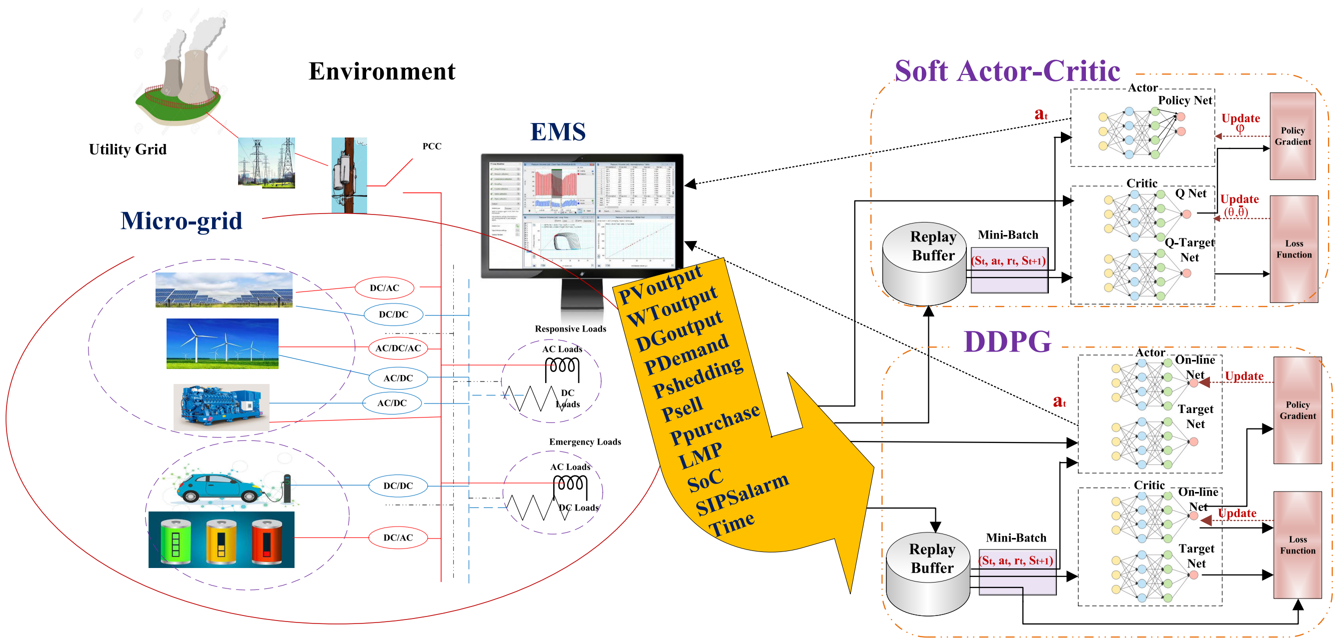


FIGURE 2. Overview of the microgrid environment and solution algorithms.

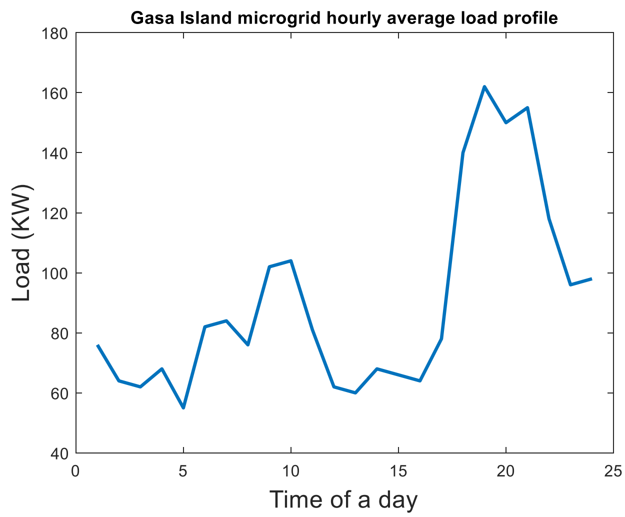


FIGURE 3. Gasa Island microgrid average daily load profile [27].

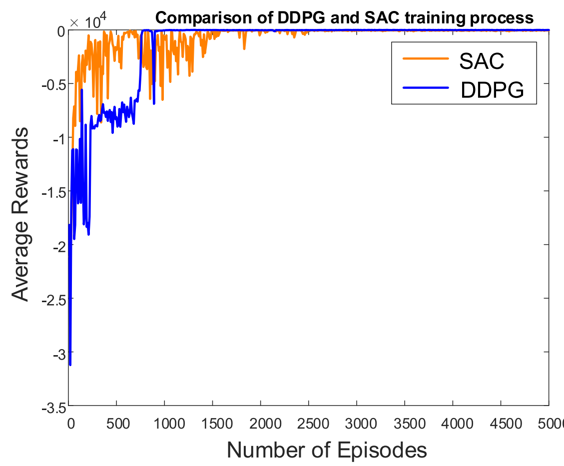
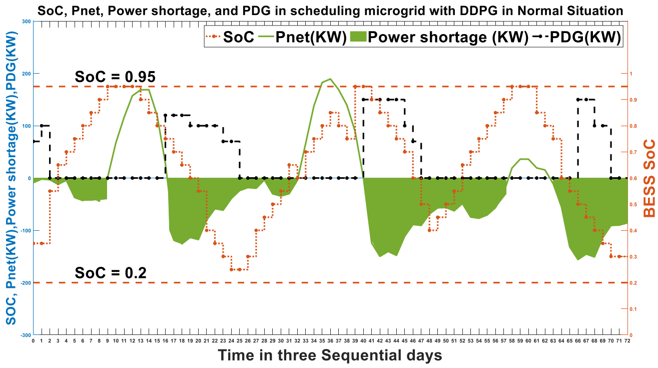


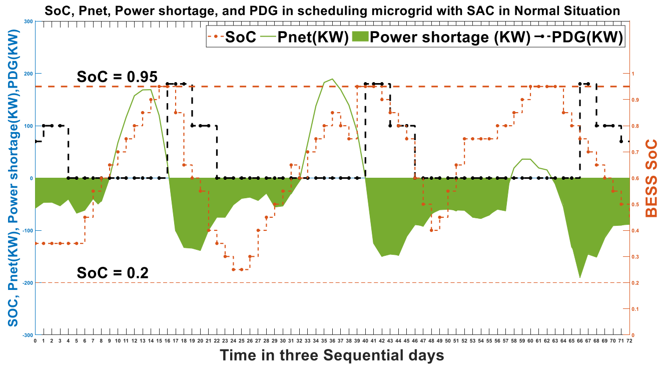
FIGURE 4. SAC and DDPG average rewards per episode.

Island's specifications to highlight the standalone microgrids community capacity in improving the power system performance, besides making a profit from participating in the electricity market. Hence, in our scenario, the microgrid acts as a prosumer trading energy in the electricity market with the capability of responding to the utility grid's SIPS alarms. As we discussed before, our microgrid environment is model-free concerning stochastic characteristics of electricity market price, loads, and RERs output power. We obtain the hourly LMP of the year 2020 from the Korean power exchange company (KPX) website [28]. The output power of WT and PV during 2020 obtained using the ninja app from [29] based on Gasa Island's installed RERs capacity. To implement demand response, we applied 15% shedding to the average daily load profile in Gasa Island. Fig.3 shows the Gasa Island average daily load profile [27]. It is estimated that Korean residential lifestyle electricity consumption can droped by 15% in contingencies [30]. The BESS charging and discharging is a continuous process in the span of [-70kW, 70kW].

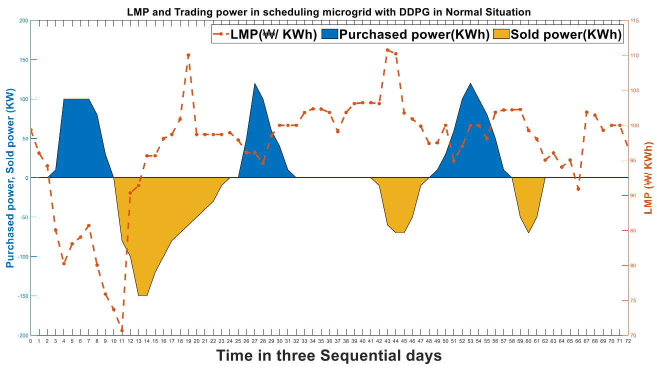
All of the hired neural networks are arranged by two fully connected hidden layers with ReLU as activation function and SGD as optimizer. The replay buffer size is 50,000, where the number of training episodes and minibatch size is 5,000 and 128, respectively. Action selection in DDPG follows the strategy of applying noise to action with mean-zero



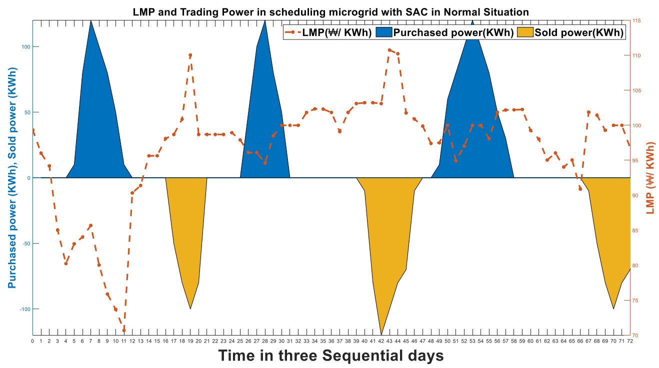
(a) DG and BESS performance with DDPG scheduling



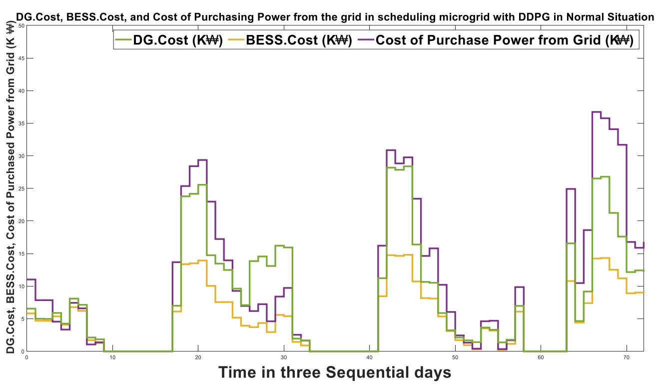
(b) DG and BESS performance with SAC scheduling



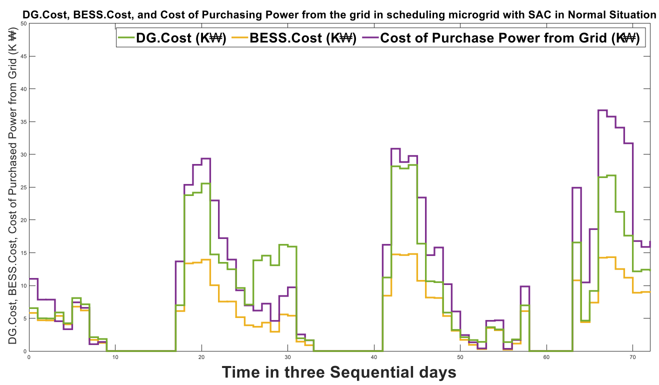
(c) Microgrid trading power with DDPG scheduling



(d) Microgrid trading power with SAC scheduling



(e) Cost of DG and BESS contribution in power provision with DDPG scheduling



(f) Cost of DG and BESS contribution in power provision with SAC scheduling

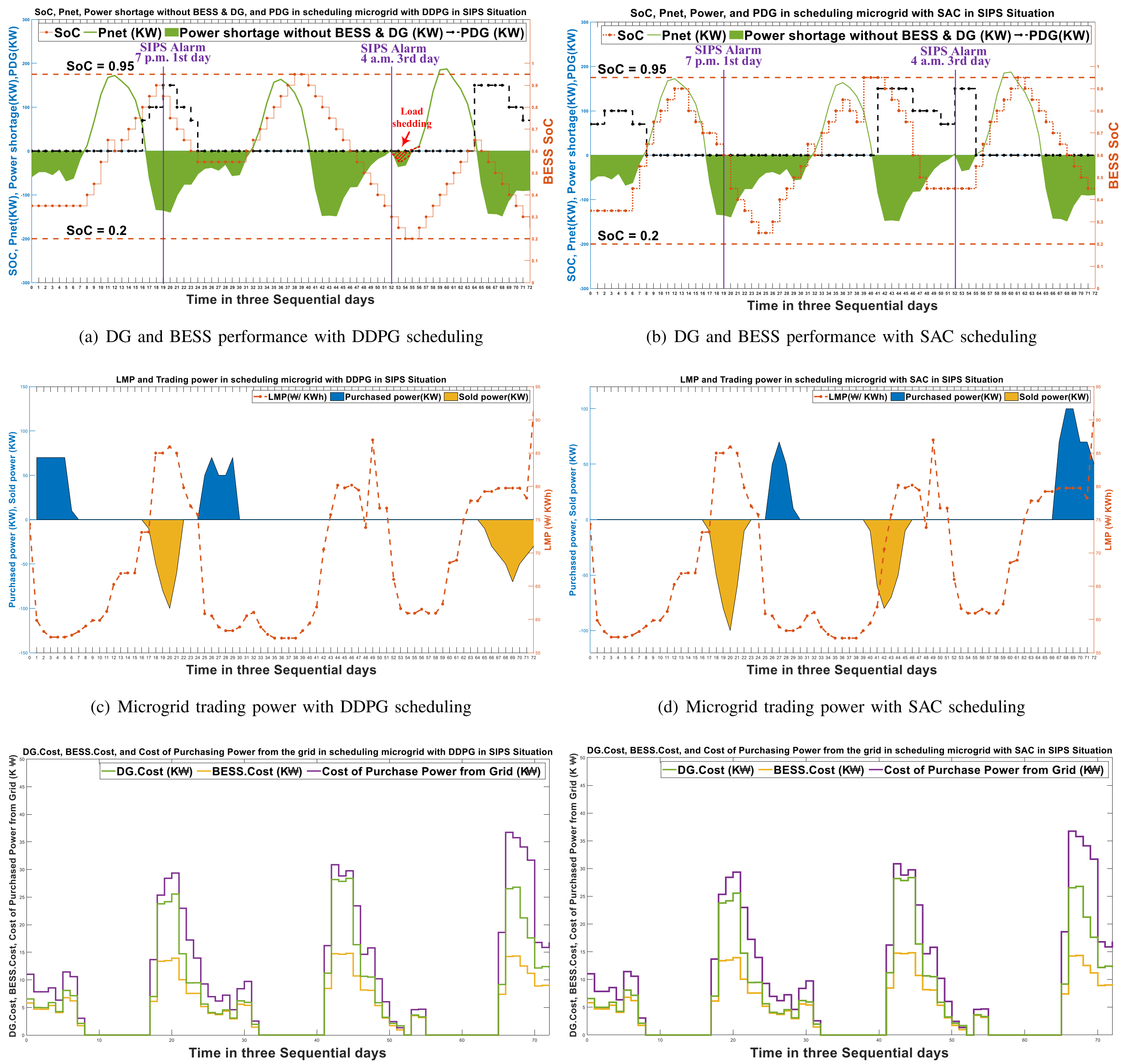
FIGURE 5. **Comparison of microgrid EMS performance with DDPG and SAC in Normal Stuation.**

Gaussian distribution that ensures exploration of a vast action state. In our scenario, we applied noise to the output of the Q-function with ε-greedy strategy according to (38). We allocated 1 to ε in 500 initial episodes to select action without noise and decayed that gradually in further episodes to select noised actions. Table 2 delineates hyperparameters of both proposed algorithms.

$$a_t = \begin{cases} Any \quad a_t, & probability\ \varepsilon \\ a_t = \mu(\mathcal{S}|\theta^{\mu}) + \mathcal{N}_t, & probability\ 1-\varepsilon \end{cases} \tag{38}$$

V. NUMERICAL RESULTS

In this section, we investigate the performance of both utilized algorithms in our microgrid environment proposal. We carried out our simulation on PC Intel(R) Core (TM) i5-10400F CPU @ 2.90GHz. The microgrid environment and solution algorithms were implemented on MATLAB Simulink (R2020b). As discussed before, Gasa Island microgrid data is used to model the microgrid. In this approach, EMS is trained by the information of Gasa Island microgrid during 2020, including RERs output power, loads, and LMP.



(a) DG and BESS performance with DDPG scheduling

(b) DG and BESS performance with SAC scheduling

(c) Microgrid trading power with DDPG scheduling

(d) Microgrid trading power with SAC scheduling

(e) Cost of DG and BESS contribution in power provision with DDPG scheduling

(f) Cost of DG and BESS contribution in power provision with SAC scheduling

FIGURE 6. **Comparison of microgrid EMS performance with DDPG and SAC in SIPS Stuation.**

We deployed data of eight months involving two months of each season for train and then evaluated the system performance with the left four months. Fig.4 indicates training performance comparison of DDPG and SAC algorithms. It can be seen from this figure that DDPG target achievement performance per number of episodes is superior to SAC. While DDPG converges to the goal in 1,000 steps, it takes 2,500 steps for SAC to converge. Fig.4 also reveals as we set the ε-greedy according to (38), the agent in the DDPG method explores to learn the environment in initial episodes, therefore obtains fewer rewards. With increasing the number of episodes, the agent will follow a greedy strategy and achieve more rewards. There is fluctuation in DDPG execution, and as we expected, the SAC performance is more stable than DDPG. Moreover, SAC provides higher rewards due to its excellent stochastic policy and action selection make its action space exploration more efficient.

To estimate the efficiency of the system, we consider two cases of microgrid operation. In Case $\mathcal{I}$, the microgrid interacts with the main grid in the normal situation, while in Case $\mathcal{II}$, the microgrid receives a SIPS signal from the utility grid. Fig.5 and Fig.6 propose our method examination for three sequential days of a month from our test data set. These figures represent DG and BESS performance, the surplus power of RERs output calculated by (16), amount of microgrid trading power with the utility grid, and cost comparison of the power delivered by DG, BESS, and purchasing from the grid with DDPG and SAC scheduling in the normal and SIPS situations, respectively. The microgrid elements performance in all chosen days proves the accuracy of scheduling the microgrid in both DDPG and SAC algorithms with a slight difference in their operations. The major suppliers of the microgrid are RERs. In the time of these resources' absence, where the BESS's SoC level is not enough to supply loads cost of power delivering by DG and utility grid determine the power supplier. BESS charges in low LMP and makes a profit by selling power to the grid in high LMP. When the microgrid receives the SIPS alarm without considering make a profit the more reliable resources among available ones are responsible to meet the utility grid requirements. If we take a look in-depth at Fig.5(b) and Fig.5(d), it reveals in the first day from 4 a.m., where LMP is low, BESS starts to charge with surplus power generated by RERs. Regarding the high LMP, BESS discharge, and microgrid sells power to the utility grid from 4 p.m. This time is the start point of working DG since RERs energy production is not enough to supply load, and BESS is enjoying making a profit with high LMP. Another interesting point about DG operation that can be seen in these figures on the third day from 6 p.m. DG injects power to the microgrid when the generation cost of DG drops to less than LMP as shown in Fig.5(f) and SoC of BESS is not enough to meet the power demand. Fig.5(a), Fig.5(c), and Fig.5(e) show the microgrid's EMS performance in the normal situation when it is trained with DDPG. According to figures 5(a) and 5(c), the microgrid purchases energy to compensate for its power shortage from the utility grid from 2 a.m. of the first day and charge BESS. DG starts to work at 5 p.m. on the first day, 4 p.m. on the second day, and 7 p.m. on the third day since the BESS's SoC level is not enough to supply loads, and the cost of power generation by DG is lower than purchasing power from the utility grid, according to Fig.5(e). Therefore, both methods have an accurate performance in hiring resources with minimum cost in energy provision.

In Case $\mathcal{II}$, microgrid response to SIPS alarm is practiced by scheduling days with fewer available RERs from our dataset test to create a severe situation. In this way, we realize how our model can manage reliable resource allocation to meet SIPS requirements. Fig.6 demonstrates results for both DDPG and SAC methods concerning the SIPS situation during three sequential days of July. We assumed the microgrid receives SIPS at 7 p.m. on the first day and 4 a.m. on the third day. In the first SIPS situation, with the DDPG learning method, BESS is the highest reliable power source in the microgrid that discharges to provide power requirements of the utility grid without considering LMP price as shown in Fig.6(a), Fig.6(c), and Fig.6(e). The load drops 15% to satisfy the power grid requirements in the other SIPS situation, where the SoC level prevents BESS from contributing to SIPS. Figures 6(b), 6(d), and 6(f) reveal SIPS situation experience with the SAC method. It is observed that in the first SIPS situation, EMS with SAC has the same approach in DDPG and BESS meets the utility grid SIPS requirement. However, SAC offers a more reliable source than load shedding by increasing DG power generation in the other SIPS situation. We can refer to SAC's stochastic policy characteristics lead to choosing the action with more rewards for this efficiency in the result.

VI. CONCLUSION

In this paper, we planned a DRL-based solution for EMS in the microgrid. We enhanced the integration of microgrids to the utility grid, from the electricity market participant to the power grid SIPS element, and upgraded the MDP of the microgrid environment based on this enhancement. Recent approaches in DRL, i.e., SAC and DDPG, which have different deterministic and stochastic policy strategies, respectively, were deployed. Moreover, actual data from Gasa Island microgrid hired to estimate the efficiency of the system. The results showed both algorithms perform well in scheduling microgrid elements to participate in the electricity market while responding to SIPS requirements. DDPG represented better time convergence per number of episodes, while SAC offered a stable training procedure. However, more computational resources for SAC to manage experience storages can overcome its longer time convergence that we considered as a future work. In the further attempt, we also will consider providing accurate reliability coefficients for the microgrid resources taking part in SIPS. Additionally, we will examine the system performance in different protection experiences and extend the environment to the multi-agent microgrid space.

REFERENCES

[1] N. Bizon and I. C. Hoarcă, "Hydrogen saving through optimized control of both fueling flows of the fuel cell hybrid power system under a variable load demand and an unknown renewable power profile," *Energy Convers. Manage.*, vol. 184, pp. 1–14, Mar. 2019.
[2] L. Tightiz, H. Yang, and M. J. Piran, "A survey on enhanced smart microgrid management system with modern wireless technology contribution," *Energies*, vol. 13, no. 9, p. 2258, May 2020.
[3] *IEEE Standard for the Specification of Microgrid Controllers*, IEEE Standard 2030.7-2017, 2018, pp. 1–43.
[4] A. Chaouachi, R. M. Kamel, R. Andoulsi, and K. Nagasaka, "Multiobjective intelligent energy management for a microgrid," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1688–1699, Apr. 2013.
[5] G. Kumar Venayagamoorthy, R. K. Sharma, P. K. Gautam, and A. Ahmadi, "Dynamic energy management system for a smart microgrid," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1643–1656, Aug. 2016.
[6] R. Palma-Behnke, C. Benavides, F. Lanas, B. Severino, L. Reyes, J. Llanos, and D. Sáez, "A microgrid energy management system based on the rolling horizon strategy," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 996–1006, Jun. 2013.
[7] F. A. Mohamed and H. N. Koivo, "System modelling and online optimal management of microgrid using mesh adaptive direct search," *Int. J. Electr. Power Energy Syst.*, vol. 32, no. 5, pp. 398–407, Jun. 2010.

[8] S. A. P. Kani, H. Nehrir, C. Colson, and C. Wang, "Real-time energy management of a stand-alone hybrid wind-microturbine energy system using particle swarm optimization," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2011, p. 1.

[9] M. S. Mahmoud, N. M. Alyazidi, and M. I. Abouheaf, "Adaptive intelligent techniques for microgrid control systems: A survey," *Int. J. Electr. Power Energy Syst.*, vol. 90, pp. 292–305, Sep. 2017.

[10] E. Kuznetsova, Y.-F. Li, C. Ruiz, E. Zio, G. Ault, and K. Bell, "Reinforcement learning for microgrid energy management," *Energy*, vol. 59, pp. 133–146, Sep. 2013.

[11] S. Kim and H. Lim, "Reinforcement learning based energy management algorithm for smart energy buildings," *Energies*, vol. 11, no. 8, p. 2010, Aug. 2018.

[12] V. François-Lavet, D. Taralla, D. Ernst, and R. Fonteneau, "Deep reinforcement learning solutions for energy microgrids management," in *Proc. Eur. Workshop Reinforcement Learn. (EWRL)*, 2016, pp. 1–7.

[13] D. Domínguez-Barbero, J. García-González, M. A. Sanz-Bobi, and E. F. Sánchez-Úbeda, "Optimising a microgrid system by deep reinforcement learning techniques," *Energies*, vol. 13, no. 11, p. 2830, Jun. 2020.

[14] Y. Ji, J. Wang, J. Xu, X. Fang, and H. Zhang, "Real-time energy management of a microgrid using deep reinforcement learning," *Energies*, vol. 12, no. 12, p. 2291, Jun. 2019.

[15] V. Bui, A. Hussain, and H. Kim, "Double deep *Q*-learning-based distributed operation of battery energy storage system considering uncertainties," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 457–469, Jan. 2020.

[16] E. Mocanu, D. C. Mocanu, P. H. Nguyen, A. Liotta, M. E. Webber, M. Gibescu, and J. G. Slootweg, "On-line building energy optimization using deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3698–3708, Jul. 2019.

[17] T. A. Nakabi and P. Toivanen, "Deep reinforcement learning for energy management in a microgrid with flexible demand," *Sustain. Energy, Grids Netw.*, vol. 25, Mar. 2021, Art. no. 100413.

[18] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with deep reinforcement learning," 2013, *arXiv:1312.5602*. [Online]. Available: http://arxiv.org/abs/1312.5602

[19] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double *Q*-learning," 2015, *arXiv:1509.06461*. [Online]. Available: http://arxiv.org/abs/1509.06461

[20] N. K. Rajalwal and D. Ghosh, "Recent trends in integrity protection of power system: A literature review," *Int. Trans. Electr. Energy Syst.*, vol. 30, no. 10, 2020, Art. no. e12523.

[21] V. Madani, D. Novosel, S. Horowitz, M. Adamiak, J. Amantegui, D. Karlsson, S. Imai, and A. Apostolov, "IEEE PSRC report on global industry experiences with system integrity protection schemes (SIPS)," *IEEE Trans. Power Del.*, vol. 25, no. 4, pp. 2143–2155, Oct. 2010.

[22] F. Luo, Z. Y. Dong, K. Meng, J. Qiu, J. Yang, and K. P. Wong, "Short-term operational planning framework for virtual power plants with high renewable penetrations," *IET Renew. Power Gener.*, vol. 10, no. 5, pp. 623–633, May 2016.

[23] M. Farrokhabadi, C. A. Cañizares, and K. Bhattacharya, "Unit commitment for isolated microgrids considering frequency control," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3270–3280, Jul. 2018.

[24] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*. [Online]. Available: http://arxiv.org/abs/1509.02971

[25] Z. Hou, K. Zhang, Y. Wan, D. Li, C. Fu, and H. Yu, "Off-policy maximum entropy reinforcement learning: Soft actor-critic with advantage weighted mixture policy(SAC-AWMP)," 2020, *arXiv:2002.02829*. [Online]. Available: http://arxiv.org/abs/2002.02829

[26] T. Haarnoja, H. Tang, P. Abbeel, and S. Levine, "Reinforcement learning with deep energy-based policies," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1352–1361.

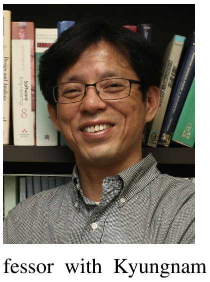
[27] M. Husein, V. B. Hau, I.-Y. Chung, W.-K. Chae, and H.-J. Lee, "Design and dynamic performance analysis of a stand-alone microgrid—A case study of Gasa Island, South Korea," *J. Electr. Eng. Technol.*, vol. 12, no. 5, pp. 1777–1788, 2017.

[28] (Mar. 2021). *Korean Electric Power Statistics Information System, System Marginal Price*. [Online]. Available: http://epsis.kpx.or.kr/epsisnew/selectEkmaSmpShdChart.do?menuId=040202

[29] (Mar. 2021). *Open Energy System Databases & Renewables.Ninja*. [Online]. Available: http://www.renewables.ninja

[30] Y.-K. Seo and W.-H. Hong, "Constructing electricity load profile and formulating load pattern for urban apartment in Korea," *Energy Buildings*, vol. 78, pp. 222–230, Aug. 2014.

LILIA TIGHTIZ (Student Member, IEEE) received the B.Sc. degree in electronic engineering from the University of Zanjan, Iran, in 2000, and the M.Sc. degree in electronic engineering from the Islamic Azad University of Ashtian, Iran, in 2015. She is currently pursuing the Ph.D. degree in computer science engineering with Sejong University, Seoul, South Korea. She has more than 15 years of experience with Tehran Electricity Distribution Company, as a Power Distribution Engineer with a history of working in the design, utilization, and maintenance of electricity distribution grid with several patents and world-class prizes in this area. She joined the Next-Generation Network Laboratory, Sejong University, in 2018. Her research interests include smart grid communication structure, IEC 61850, the IoT, microgrid energy management systems, and deep reinforcement learning.



HYOSIK YANG (Member, IEEE) received the B.E. degree in information and communication engineering from Myongji University, Seoul, South Korea, in 1998, and the M.S. and Ph.D. degrees in electrical engineering from Arizona State University, Tempe, AZ, USA, in 2000 and 2005, respectively. He is currently a Professor with the Department of Computer Science and Engineering, Sejong University, Seoul. Before he joined Sejong University, he was an Assistant Professor with Kyungnam University, Changwon, South Korea. He joined Sejong University, in 2006. He has been serving as a Faculty Research Associate with Arizona State University, since 2005. His research interests include wavelength-division multiplexing (WDM) all-optical networks, mobile *ad-hoc* networks, smart grid, especially on IEC 61850 and communication architecture, and smart city. He is a member of IEC TC 57 WG10, and IEC SyC Smart City and Smart Energy.