Multi-Agent Reinforcement Learning for Decentralized Resilient Secondary Control of Energy Storage Systems Against DoS Attacks

Pengcheng Chen, Shichao Liu, *Senior Member, IEEE*, Bo Chen, *Member, IEEE*, and Li Yu, *Member, IEEE*

***Abstract*—While distributed secondary controllers have been studied for multiple energy storage systems in islanded microgrids, information infrastructure has to be added for the extensive information transmission among these secondary controllers and the additional communication among distributed controllers is costly and increases the vulnerability surface to cyberattacks. In this work, a data-driven decentralized secondary control scheme is proposed for multiple heterogeneous battery energy storage systems (BESSs). The proposed secondary control scheme can achieve frequency regulation and the state-of-charge (SoC) balancing simultaneously for BESSs without requiring accurate BESS models. This scheme leverages an asynchronous advantage actor-critic (A3C) based multi-agent deep reinforcement learning (MA-DRL) algorithm where the centralized off-line learning with shared convolutional neural networks (CNN) is designed to maximize global rewards and ensure the performance of the entire system and a decentralized online execution mechanism is applied to each BESS. Furthermore, in view of possible denial-of-service (DoS) attack on local communication networks used for signal transfer between secondary controllers and remote sensors, a signal-to-interference-plus-noise ratio (SINR)-based dynamic and proactive event-triggered communication mechanism is proposed to alleviate the impact of DoS attacks and reduce the occupation of communication resources. Simulation results on a four-bus multiple BESS system show that the proposed decentralized secondary controller can achieve simultaneous frequency regulation and SoC balancing. Comparison results with other event-triggered mechanisms and MA-DRL algorithms show the A3C based MA-DRL algorithm with CNN can obtain a comparatively optimal policy through training and the designed event-triggered strategy can dynamically adapt the release frequency based on real-time SINR and significantly reduce the occupied network bandwidth and packet loss rate (PER) induced by DoS attacks.**

***Index Terms*—Multi-agent deep reinforcement learning, microgrid, secondary frequency control, DoS attacks, battery energy storage systems.**

Manuscript received August 19, 2021; revised November 15, 2021; accepted January 6, 2022. Date of publication January 11, 2022; date of current version April 22, 2022. This work was supported in part by the NSERC Discovery Grant; in part by the National Natural Science Funds of China under Grant 62073292; and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LR20F030004. Paper no. TSG-01327-2021. *(Corresponding author: Shichao Liu.)*

Pengcheng Chen and Shichao Liu are with the Department of Electronics, Carleton University, Ottawa, ON K1S5B6, Canada (e-mail: pengchengchen@cmail.carleton.ca; shichaoliu@cunet.carleton.ca).

Bo Chen and Li Yu are with the Department of Automation and the Institute of Cyberspace Security, Zhejiang University of Technology, Hangzhou 310023, China (e-mail: bchen@aliyun.com; lyu@zjut.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSG.2022.3142087.

Digital Object Identifier 10.1109/TSG.2022.3142087

I. Introduction

DURING the past decades, distributed multiple battery energy storage systems (BESSs) have played a significant role in microgrid control and operation due to the increasing penetration of renewable energy sources and electrical vehicles [1], [2]. The inclusion of BESSs in microgrid can improve the power quality and network stability by properly charging/discharging to balance the intermittent power generation and time-varying demand [3]. Although the primary droop control can fast stabilize the microgrid, it causes the deviation of the steady-state frequency in BESSs from the nominal frequency [4]. Thus, secondary control is critical to restore the microgrid voltage and frequency to their nominal values. When multiple heterogeneous BESSs participate in secondary frequency control in microgrid, the energy levels or the state of charge (SoC) levels of these BESSs should be coordinated properly to guarantee that the BESSs will not lack of energy or be overloaded. Balancing SoC levels of BESSs can extend the life cycle of the BESSs while having the abilities of power quality enhancement and peak shaving/shifting [5], but it will also cause fluctuations in active power and in turn degrade the frequency regulation. Therefore, it is obligatory to coordinate the SFC [6] and SoC balancing [7] issues simultaneously.

On the other hand, cyber attacks, e.g., denial-of-service (DoS) attacks [8] and false data injection (FDI) attacks [9], targeting on control systems in power systems have occurred with an increasing frequency, such as the 2020 Mumbai power outage and the 2016 Ukrainian power blackout [10]. The DoS attackers block the communication channel, so that the network data cannot be successfully transmitted, which destroys the validity of the data. What's worse, the generation portfolio cannot receive effective control signals, so that it will further deteriorate the performance of the BESSs [11]. Therefore, how to design an effective control strategy for BESSs to deal with SFC and SoC balancing problems subject to DoS attacks on local communication networks between secondary controllers and primary controllers is still an intractable and significant problem.

Conventional control strategies for BESSs are mainly model-based [5] and their system performance highly depends on the accuracy of modeling. However, due to BESS heating or external factors, it is difficult to model the electrical components of BESSs accurately during operation [12]. The MA-DRL algorithm is a powerful data-driven approach, which

can overcome the non-stationary problem through the role of each individual agent [13], [14] compared with the traditional reinforcement learning method. An attention enabled MA-DRL algorithm was proposed in [15] to deal with the decentralized Volt-VAR control problem in the distribution network, where the entire power distribution system was divided into some sub-networks based on unsupervised clustering. In [16], an MA-DRL algorithm was designed to manage energy for multiple charging stations in BESSs with the solar photovoltaic energy. In order to coordinate reactive and active power of photovoltaic energy with high penetration in BESSs, a soft MA-DRL algorithm with centralized learning and decentralized execution structure was devised in [17] to reduce communication between each controller. Considering the challenges of fast demand responses and expansion of energy systems, a multi-agent deep deterministic policy gradient approach was proposed in [18] to cope with the autonomous voltage control problem, which could train control agents from scratch and obtain system operation law by learning from data produced by the energy system. In order to enhance the resilient of power systems, an MA-DRL based hybrid soft actor-critic algorithm was investigated in [19] to deal with the voltage violations because of outages of transmission lines when wind storms happened. However, the simultaneous SFC and SoC balancing coordination is not touched in the above work and the cyberattack issue is not their focus.

On the other hand, the event-triggered (ET) mechanism is used to weaken the impact of communication delay [20] and DoS attacks [21] because it can reduce the number of communications while ensuring system performance, thereby reducing the occupation of network bandwidth. However, the traditional event-triggered threshold is constant [20], [22], resulting in the constant release frequency regardless of system operating conditions and variation in the external environment. There were some dynamic event-triggered mechanisms proposed for power systems recently. In [23], a resilient event-triggered communication strategy was proposed for load frequency control to keep the frequency stable under DoS attacks, where an uncertain item related to the DoS attack was introduced into the triggered condition. Considering that there are different droop control characteristics between the BESSs and distributed generators, a unified dynamic event-triggered method combined with the distributed active power sharing control was proposed in [24] to adjust different charging power and balance the SoC. In order to achieve optimal active power sharing and frequency/voltage restoration in distributed stochastic secondary control with limited bandwidth constraints and noise interferences, an event-triggered communication strategy was proposed in [25] to reduce the influence of the external environment. An adaptive event-triggered was proposed in [26] to deal with the load frequency control problem in multi-area power systems with communication delay, where its threshold was determined by historical system state or observed system state. However, to the best knowledge of authors, most existing dynamic event-triggered mechanisms are self-adaptive to monitor the state of the system [20]–[22], [25], [26]. Corresponding measures will be triggered only after the state of the system is affected by external

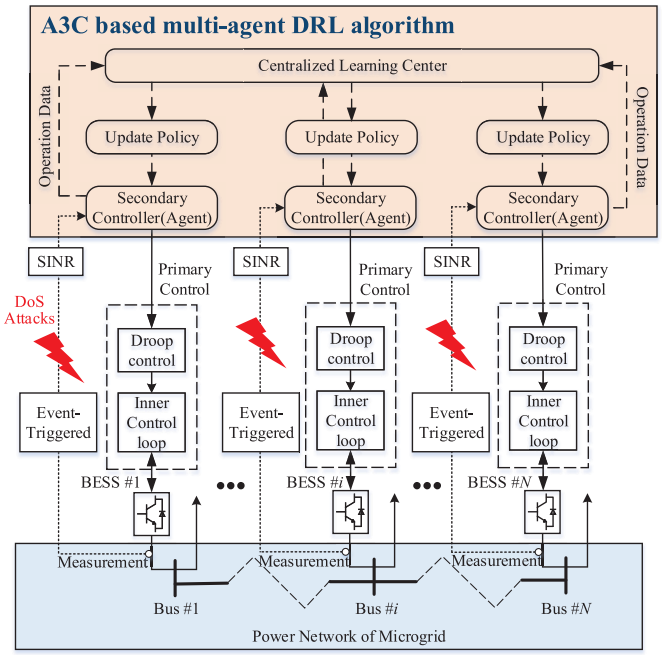

Fig. 1. The designed system framework under DoS attacks for BESSs.

malignant factors, which are relatively passive. Moreover, the existing researches on DoS attacks have relatively strong assumptions, that is, DoS attacks have a certain statistical distribution [27], [28] or the attack's dwell-time [21] or period is fixed [23], [29]. However, in reality, DoS attacks do not follow any given information, so the countermeasures can only be designed from the source of the attacks. It is necessary to develop the event-triggered mechanism that can substantially reduce the impact of DoS attacks, and there are few such studies at present.

This work presents a data-driven decentralized secondary control scheme proposed for multiple heterogeneous BESSs in islanded microgrids as shown in Fig. 1, where measurements are sent from remote terminal units (RTUs) via the supervisory control and data acquisition (SCADA) systems to secondary controllers. The contributions and novelties are summarized as follows.

- The proposed secondary control follows a decentralized structure and therefore avoids the extensive real-time data exchange among secondary controllers when distributed structures are used. To the best knowledge of the authors, this is the first attempt in developing a data-driven decentralized secondary controllers for microgrids with multiple heterogeneous BESSs.
- The proposed secondary control scheme achieves secondary frequency regulation and the state-of-charge (SoC) balancing simultaneously for multiple heterogeneous BESSs without requiring accurate BESS models.
- There are two control loops in the proposed secondary control scheme. The inner loop is a signal-to-interference-plus-noise ratio (SINR)-based dynamic and proactive event-triggered communication mechanism to alleviate the impact of DoS attacks on the local communication network for transferring measurements between secondary controllers and

remote sensors. The outer loop is the multi-agent deep reinforcement learning framework and its environment include the inner loop variables such as SINR, event-triggered mechanism, DoS attacks and BESSs. In the outer loop, the agents receive observation given by the inner loop and reward to update value functions and policy functions and offer action/control signal to realize simultaneous secondary frequency control (SFC) and the state-of-charge (SoC) balancing.

- Simulation results indicate, compared to a variety of event-triggered mechanisms and MA-DRL algorithms, the A3C based MA-DRL algorithm with CNN can obtain a comparatively optimal policy through training and the designed event-triggered strategy can dynamically adapt the release frequency based on real-time SINR and significantly reduce the occupied network bandwidth and packet loss rate (PER) induced by DoS attacks.

The rest of the paper is presented as follows: the SFC and SoC balancing model with SINR-based event-triggered strategy under DoS attacks of the multiple BESSs is introduced in Section II. Section III shows the framework of the proposed method and an MA-DRL model for optimizing BESSs performance while an A3C based MA-DRL algorithm is designed to solve these problems. Section IV shows the simulation results of the training process for verifying the effectiveness of the proposed method. Conclusions are presented in Section V.

II. Systems Model and Problem Formulation

In this section, we give the SFC and SoC balancing model combined with SINR-based event-triggered mechanism under DoS attacks for BESSs. In Section II-A and Section II-B, we present the system model of SFC and SoC balancing, respectively. An SINR-based event-triggered strategy is designed in Section II-C under DoS attacks.

A. Secondary Frequency Control

In BESSs, the operation of synchronous generators could be emulated by droop control of power inverters, which is employed to accommodate the voltage V_i and system frequency ω_i only based on local measurement of reactive power Q_i and real power P_i, respectively. It shows the typical droop control as follows [30], [31]:

$$\omega_i(t) = \omega_i^{\text{nom}} - K_i^w P_i(t) \tag{1}$$

$$V_i(t) = V_i^{\text{nom}} - K_i^V Q_i(t),\ \ i = 1, \ldots, N \tag{2}$$

where V_i^{nom} and ω_i^{nom} are the nominal value of voltage and frequency, respectively. The droop parameters are K_i^w and K_i^V, which are determined by $K_i^w = \Delta\omega/P_i^{\max}$ and $K_i^V = \Delta V/Q_i^{\max}$. It should be noted that the ω_i is represented as angular frequency with unit rad/s and the V_i is represented as voltage amplitude with unit U.

In this paper, we consider the impact of droop control during the charging/discharging process. Since the charging/discharging voltage is basically constant [2], it is equal to the grid voltage. Besides, the active power and reactive power are perpendicular to each other in terms of phase angle, so there is no component influence. Therefore, due to the different acting time period and time scale and focusing on the medium-frequency disturbance [5], the dynamics of the inner control loops, phase-locked-loop, filters, and the reactive power sharing/control are not included in the this work.

If the charging/discharging power is variable, the system frequency will not remain stable due to the influence of the droop parameter in the primary droop control. Thus, it is necessary to introduce the SFC to restore the frequency back to the targeted operating point [32]:

$$\omega_i = \omega_i^{\text{nom}} - K_i^w P_i + u_i^\omega \tag{3}$$

Because the proportional-integral (PI) controller has the advantages of fastness and elimination of steady-state errors, this paper adopts the frequency controller as shown below:

$$u_i^\omega = K_p^i \Delta w_i + K_i^i \frac{\Delta w_i}{s} \tag{4}$$

where $\Delta w_i = w_i - w_{\text{ref}}$. Although PI controller has relatively good control performance, it is difficult to quantitatively design gain parameters according to the required performance of the system. In this paper, we apply the A3C based MA-DRL algorithm to train proportional parameter K_p^i and integral parameter K_i^i to make the system frequency reach consensus as soon as possible.

B. SoC Balancing

In this paper, the Coulomb counting method [33] is utilized to represent the SoC in each BESS as follows

$$S_i(t) = S_i(0) - \frac{\rho_i}{Q_i} \int I_i(t)dt,\ \ i = 1, \ldots, N \tag{5}$$

where S_i represents the SoC status of ith BESS, $S_i(0)$ denotes the initial battery status, Q_i is the BESS capacity, I_i represent the battery current of ith BESS, and the ρ_i is denoted as follows

$$\rho_i = \begin{cases} 1, & \text{discharging} \\ \eta_i, & \text{charging} \end{cases} \tag{6}$$

where η_i is defined as Coulombic parameter. We denote V_i as the output voltage, then the power output P_i in ith BESS can be represented as

$$P_i = V_i I_i. \tag{7}$$

During the charging/discharging process, it is assumed that the output voltage V_i of BESS basically does not fluctuate. Then, it yields

$$S_i(t) = S_i(0) - \frac{\rho_i}{Q_i V_i} \int P_i(t)dt. \tag{8}$$

In the power system, grid operators usually give the overall active power requirement value P_{ref} according to specific tasks for all BESS, so the sum of the active power output of each BESS must meet the following constraints

$$P = \sum_{i=1}^{N} P_i = P_{\text{ref}}. \tag{9}$$

In order to meet the requirements for overall active power (9), a simple control method is to output the average

power of each BESS, i.e., $P_i = P_{\text{ref}}/N$. However, since the capacity of each BESS is different in (5), $S_i(t)$ will diverge even if there is the same initial $S_i(0)$ for all BESSs. Besides, during the absorbing or supplying power process, for the purpose of keeping the overall maximum capacity of all BESSs while prolonging battery life, it is not desirable for any BESS to be disconnected from the power systems in advance due to full charge or exhaustion. Therefore, it is reasonable to control the charging/discharging power, according to the capacitance of electrical appliances and the initial $S_i(0)$, in order to achieve the convergence of $S_i(t)$ for all BESSs. Considering the influence of power changes on frequency (3), it is necessary to address SFC and SoC balancing issues simultaneously.

In this paper, we propose an SoC-based active power sharing method, which is also found in [2]. Being different from the active power sharing achieved via the adaptive droop controller in [2], the MA-DRL algorithm is employed to allocate active power reasonably to sustain the goal of SoC balancing with different initial SoC values. In detail, BESSs with lower SoC values are allocated a higher charging power while the ones with higher power are allocated lower charging power to achieve SoC balancing while meeting the requirements for overall active power (9) in the steady state. After reaching the average value, in order to continue maintaining the SoC balancing, the increment value of SoC in each BESS has to be consistent. According to Coulomb counting method in (8), we can obtain

$$\frac{P_i}{P_j} = \frac{\rho_j Q_i V_i}{\rho_i Q_j V_j}. \tag{10}$$

For SoC balancing, the control signal can be chosen as the active power of each BESS as follows:

$$u_i^S = P_i \tag{11}$$

which will be trained by the A3C based MA-DRL algorithm proposed in this paper.

C. SINR-Based Event-Triggered Strategy Under DoS Attacks

Due to the importance of the power system to economic development and social stability, it often becomes a key target of cyber attacks. DoS attacks prevent the controller from obtaining effective information by blocking the communication channel, thereby destroying the performance of the power system. Usually, a mass of puppet hosts are operated by DoS attackers to send a large amount of invalid data to the communication channel of the key electrical nodes, which results in the network bandwidth or CPU resources of the attacked host being exhausted [34]. It is assumed in this paper that the DoS attackers only block the channels from sensor of microgrid to secondary controllers as shown in Fig. 1. SINR is the ratio of effective signal power to external interference and noise power, which can be used to measure the congestion extent of the communication channel under DoS attacks and obtain the quantitative relationship between DoS attacks frequency and SINR [35]. It is assumed that the transmission channel of each BESS can be equivalent to one. The SINR of the communication channel in ith BESS can be expressed as follows:

$$\text{SINR}_i(t) = \frac{T_i}{\sum_{j=1}^{N} T_j + J + (\sigma_i)^2} \tag{12}$$

where the T_i represents the transmission power of communication channel in ith BESS, J is transmission power which is used to block all channels in BESSs by DoS attackers, and $(\sigma_i)^2$ represents the background noise in ith BESS, which can be different in each BESS. According to digital communication theory [36], the inherent connection between SINR and packet loss rate (PER) induced by DoS attacks can be expressed as follows:

$$\text{PER}_i(t) = f(\text{SINR}_i(t)) \tag{13}$$

where $f(\cdot)$ is a monotonically decreasing function, which is determined by modulation and characteristic mechanisms.

In this work, the DoS attacks are not assumed to follow specific statistical distributions, fixed dwell-time, or period. SINR is a more realistic way to characterize DoS attacks, since the result of the attack is considered as packet loss in this work. It is noted that there are three factors that can affect the SINR in ith BESS, i.e., the transmission power of other BESSs, the blocking power of the attackers and the background noise. Intuitively, if the attacker's blocking power is weaker, other BESSs transmit less simultaneously, and the background noise is weaker, then the SINR of this channel is greater, which means the greater the possibility of successful transmission. It is difficult and impractical for the defender to reduce the possibility of being attacked by reducing the blocking power of attackers and background noise. Thus, it is necessary to design a communication mechanism that guarantees system performance while reducing the releases number of all BESSs, which can fundamentally alleviate the impact of DoS attacks. In this paper, we devise the following SINR-based dynamic event-triggered communication mechanism for SFC and SoC balancing:

$$\begin{aligned}&\left[y_i\left(t_{k,i}h\right) - y_{avg}\left(t_{k,i}h\right)\right]^T\left[y_i\left(t_{k,i}h\right) - y_{avg}\left(t_{k,i}h\right)\right]\\ &\quad > \sigma_i(t) y_i^T\left(t_{k,i}h\right) y_i\left(t_{k,i}h\right)\end{aligned} \tag{14}$$

$$\sigma_i(t) = \sigma_{im} - \sigma_{im}\tanh(\text{SINR}_i(t)) \tag{15}$$

where the $y_i(t_{k,i}h)$ could be $S_i(t)$ or $\omega_i(t)$, and $y_{avg}(t_{k,i}h) = \sum_1^N S_i(t)/N$ or $y_{avg}(t_{k,i}h) = \sum_1^N \omega_i(t)/N$, respectively. It is noted that y_{avg} needs global information to calculate it, which can be realized by collaborative working mode of agents and the centralized learning mechanism. Besides, it is good for consensus of SFC and SoC balancing.

The event-triggered mechanism proposed in this paper is to proactively adjust the data release process to reduce the occupation of communication resources while ensuring system performance, thereby reducing the possibility of being DoS attacks. DoS attacker often maliciously employs a large number of puppet computers to send a large amount of useless data to the communication channel between sensor and secondary controller, which results in the network bandwidth of the attacked host being exhausted and signal-to-interference-plus-noise ratio (SINR) signal being smaller. Consequently,

secondary controller cannot receive effective measurement signals. We use SINR signal of communication network which could reflect the congestion level of the network, to describe PER of DoS attacks. An SINR-based dynamic event-triggered mechanism is then proposed to schedule the transmission process to deal with DoS attacks. Furthermore, the dynamic threshold $\sigma_i(t)$ can be adjusted automatically according to $\text{SINR}_i(t)$. Intuitively, if $\text{SINR}_i(t)$ becomes larger, which means $\text{PER}_i(t)$ becomes smaller and the state of other BESSs is not necessary to be released under the effect of event-triggered condition (14). In such an unobstructed network environment, the threshold $\sigma_i(t)$ will be reduced based on dynamic adjustment mechanism (15) to increase the release frequency. Otherwise, if $\text{SINR}_i(t)$ becomes smaller, which means $\text{PER}_i(t)$ becomes larger and other BESSs urgently need communication resources to ensure system performance. In such the congested network environment, the threshold $\sigma_i(t)$ will be increased based on dynamic adjustment mechanism (15) to decrease the release frequency, so that it will reduce unnecessary data transmission to decrease the level of network congestion, further contributing to the mitigation of packet loss caused by DoS attacks. Obviously, $\sigma_i(t) \leq 2\sigma_{im}$, setting the upper bound of the threshold of event-triggered condition can ensure system stability.

Zeno phenomenon means that the interval between two adjacent released instants of event-triggered mechanism must be strictly greater than zero, which guarantees no occurrence of unlimited release times within a limited time interval. We use the sampling-based method to design the event-triggered strategy, so the minimum interval between two adjacent released instant is the sampling period h. Because h is larger than zero, the Zeno phenomenon will not occur in this method.

Compared with other dynamic event-triggered mechanisms based on the system state [26], the event-triggered mechanism designed in this paper is more proactive in reducing the impact of DoS attacks. Because, the DoS attacks may cause the system state to oscillate or diverge. If the countermeasures are designed according to the deteriorating system status, it means that the impact of DoS attacks has already occurred, so the countermeasures are relatively passive. Thus, the way to adapt the released frequency by detecting $\text{SINR}_i(t)$ proposed in this paper is more proactive.

III. Decentralized Secondary Frequency Control and SoC Balancing via A3C Based MA-DRL

In this section, we propose an A3C based MA-DRL algorithm to achieve the SFC and SoC balancing simultaneously with an event-triggered communication mechanism subject to DoS attacks on the local communication network between secondary controller and primary controller. In Section III-A, these two issues are formulated into the Markov Games. Then, an A3C based MA-DRL algorithm is given with a centralized learning and decentralized execution mechanism in Section III-B.

A. BESSs in Cooperative Markov Games

Markov games are also seen as Markov decision processes (MDP) where a common environment can be interacted by multiple agents locally [37]. Due to the complexity and variability of BESS model parameters, it is difficult to accurately model, so Markov game with sequential decisions and model-free features can show superiorities. The SFC and SoC balancing coordination in BESSs is described by a cooperative Markov games with N agents in this paper, which could be defined through a tuple of $(\mathcal{S}, [\mathcal{O}_i]_N, [\mathcal{A}_i]_N, [\mathcal{R}_i]_N, \rho, \gamma)$. The detailed description of state space, observation space, action space, and reward function are devised as follows.

- State space $\mathcal{S}$ represents all possible states in the common environment. The state $s \in \mathcal{S}$ of Markov games is designed as a vector $s = (\mathbf{w}, \mathbf{E}, t)$, where $\mathbf{w}$ is the vector of system frequency $\omega_i, \forall i \in \mathcal{N}$, and $\mathbf{E}$ is the vector of SoC state $S_i, \forall i \in \mathcal{N}$. t means the step of each episode.
- Observation space $[\mathcal{O}_i]_N$ represents the local observation data of each agent, which is determined by local measurements. In order to deal with the SFC and SoC balancing problems simultaneously, $o_i \in \mathcal{O}_i$ is designed as $(\mathbf{w}_i, \mathbf{E}_i)$, where the $\mathbf{w}_i, \mathbf{E}_i$ are the vector of system frequency ω_i and SoC state S_i in ith BESS, respectively.
- Action space $[\mathcal{A}_i]_N$ is designed as the PI parameters K_p^i, K_i^i for SFC and active power P_i for SoC balancing issues, i.e., $\mathcal{A}_i = \{K_p^i, K_i^i, P_i\}$ in ith BESS.
- Reward function $\mathcal{R}_i : \mathcal{S} \times \mathcal{A}_i \mapsto \mathbb{R}$. Different from the traditional RL algorithm, the reward function is defined the function of historical observations. Since each agent works in a collaborative mode, all observations could be used in the design of the reward function, which is conducive to achieving consensus. In order to achieve the goal of SoC balancing and stabilizing the system frequency as quickly as possible, the reward function is designed as follows:

$$r_i(t) = r_i^s(t) + r_i^\omega(t) \tag{16}$$

where

$$r_i^s(t) = K_{i1}\big(S_i(t) - S_{\text{avg}}(t)\big)^2 + K_{i2}(S_i(t) - S_{\text{tar}})^2$$
$$r_i^\omega(t) = K_{i3}(\omega_i(t) - \omega_i(t-1))^2 + K_{i4}(\omega_i(t) - \omega_{\text{tar}})^2$$

with $K_{i1}, K_{i2}, K_{i3}, K_{i4}$ are the weight matrices used to balance the importance of each item and $S_{\text{avg}}(t) = \sum_1^N S_i(t)/N$ can be obtained by coordinator. Specifically, the first item of $r_i^s(t)$ is designed to achieve consensus of SoC balancing, and the first item of $r_i^\omega(t)$ is designed to stabilize the system frequency as quickly as possible. Besides, the second item of $r_i^s(t)$ and $r_i^\omega(t)$ is designed to guide the direction of the training target for SFC and SoC balancing issues.

- Discount factor $\gamma \in (0, 1]$ is used to determine the importance of future rewards and learning rate $\rho \in (0, 1]$ can affect the convergence rate and exploration range, which will be given in the simulation part.

B. A3C-Based Multi-Agent Deep Reinforcement Learning

In this paper, the A3C based MA-DRL algorithm is used to cope with the Markov games problem in the multi-agent environment. In the actor-critic (AC) framework, the deep neural network (DNN) is used to approximate the actor,

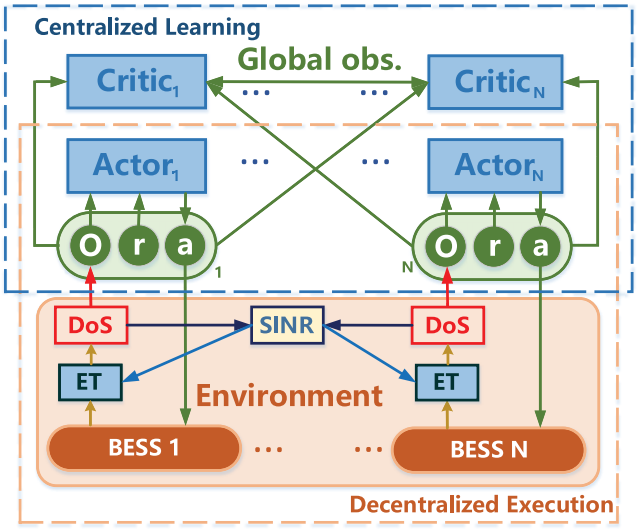

Fig. 2. The framework of multi-agent AC learning system for BESSs.

which updates policy π generates action using observation. Meanwhile, DNN also approximates the critic, which assesses the policy through V^{π} or Q^{π} to improve system performance.

In multi-agent environments, it is not suitable to utilize the conventional RL algorithm to each agent individually, because the environment is non-stationary based on the viewpoint of each agent. In this paper, the multi-agent actor-critic framework [38] is exploited to deal with the inherent non-stationary issues in multi-agent environments, and the framework of multi-agent AC learning system is displayed in Fig. 2.

In this framework, each agent possesses a local actor and a critic, where the critic is authorized to utilize global information, including actions of all actors and observations, to establish its own assessment of global environment features during the training process. As shown in Fig. 2, we set the event-triggered mechanism, SINR and DoS attacks in communication channels and BESSs as the environment for the outer MADRL loop, which means the environment has the ability to balance the occupation of network bandwidth and release data to ensure system performance. Due to the trial-and-error characteristics of deep reinforcement learning, the environment itself has the ability to resist DoS attacks caused by network bandwidth exhausted, which is meaningful for practical applications. Since the multi-agent cooperative mode is considered in this paper, we train each local actor with the corresponding critic using the information of other actors, which is the basis for achieving the requirements of overall active power and obtaining global information in the reward function and event-triggered mechanism. In order to obtain the optimal policy of the overall multi-agent learning system, the centralized training and decentralized execution mechanism is employed in this paper, which combines the superiorities of independent learning and joint learning, but it is better than them. The local actor will use the trained policy to generate the control signal in a decentralized way after the training is completed.

The asynchronous version of AC, i.e., A3C, is an on-policy algorithm combined with multi-thread training characteristics [39], which obtains the estimation of optimal policy according to the trained experience of multiple thread and has the advantages of the AC framework. For the experience replay mechanism of the off-policy algorithm, e.g., DQN, its data needs to be generated by an older policy. For the A3C algorithm, there is a large amount of time-independent data for training at each episode in each agent, and then the parameters of the global shared network are updated asynchronously. Compared with the off-policy algorithms, A3C can save computing resources and training time. CNN shares the parameters of the convolution kernel, which is used to extract features. In general, data characteristics are likely to exist in more than one place, such as "data characteristics attacked by DoS", which should be reflected in the overall data, so it is reasonable and natural for the deep learning method to share the convolution kernel. The parameter sharing mechanism and sparse connection of CNN can greatly reduce the number of network parameters. In this way, we can train a better model with fewer parameters, and can effectively avoid overfitting. Moreover, due to the parameter sharing of the convolution kernel, even if the data has some repetitiveness, the characteristics can be identified, which is called "translation invariance". Due to the parameter sharing and sparse connection of the CNN, it has the advantage of fewer parameters to be trained and it is applied to the A3C algorithm to estimate actor and critic in this paper. Different thread exploration policies are likely to be different, so the training time is linearly inversely proportional to the number of threads.

In the MA-DRL algorithm, we design multiple local CNN networks and a global CNN network in the thread of each agent. Each network has its own set of parameters, which means that they independently interact with their own copy of the environment and optimize the global neural network through asynchronous gradient descent. Therefore, this optimization mode can break the time-dependent of data generated by MDP. The excellence of this method is that the experience of each CNN is not related to the experience of other CNN, so the global experience available for training is more diverse and multiple parallel threads can contribute to exploration. In the A3C framework, the CNN is used to approximate the actor, which updates policy function π to generates action using observation. Meanwhile, CNN also approximates the critic which assesses the policy through value function to improve system performance. This algorithm framework can efficiently explore the environment and obtain a relatively optimal policy by training actors and critics to interact with each other.

Let $R_i(t)$ be the discount-accumulated reward in ith agent at step t as follows:

$$R_i(t) = \sum_{m=0}^{k-1} \gamma^m r_i(t+m) + \gamma^k V_i^{\pi_i}(o(t+k); \theta_{iv}) \qquad (17)$$

where $V_i^{\pi_i}(o(t+k); \theta_{iv})$ represents the value function under the all observation $o(t+k)$ in common environment, the reward function $r_i(t+m)$ is defined in (16), and $k \in (0, t_{\max}]$, $t_{\max}$ is the maximum step before updating. The training goal of cooperative multi-agent learning algorithm is to optimize the discount-accumulated reward of each agent for all BESSs to

maximize the global reward as follows:

$$\underset{\theta_{iv},\forall i\in\mathcal{N}}{\text{Maximize}}\mathbb{E}\left[\sum_{i=1}^{\mathcal{N}}\sum_{m=0}^{k-1}\gamma^m r_i(t+m)+\gamma^k V_i^{\pi_i}(o(t+k);\theta_{iv})\right]. \tag{18}$$

It means that SoC balancing issue can be achieved and the effect of active power on frequency can be eliminated as quickly as possible in all BESSs. Different from the value-based RL algorithm, e.g., SARSA, this work utilizes advantage function value $A_i(a_i(t), o(t))$ rather than the $Q_i(a_i(t), o(t))$ value. The role of the advantage function is to decrease the variance of $\mathbb{E}[R_i(t)]$, which is achieved through subtracting the estimated function of the state $b_i(o(t))$, also called the baseline, from the discounted-accumulated reward $R_i(t)$. It is convenient to apply the estimated value function $V_i^{\pi_i}(o(t))$ to approximate the baseline $b_i(o(t))$ and the policy gradient can be measured by $R_i(t)-b_i(o(t))$. Because $Q_i(a_i(t), o(t))$ can be estimated by $R_i(t)$ and $V_i^{\pi_i}(o(t))$ can be estimated by $b_i(o(t))$, the advantage function is chosen as $A_i(a_i(t), o(t)) = R_i(t) - V_i^{\pi_i}(o(t))$, which can be employed to approximate the advantage of action $a_i(t)$ with observation $o(t)$. The specific algorithm details for SFC and SoC balancing issues with event-triggered mechanism in BESSs are shown in Algorithm 1.

In asynchronous actor-critic framework, we regard baseline $b_i(o(t))$ and policy π_i as the output of critic and actor. The advantage function $A_i(a_i(t), o(t); \theta_i, \theta_{iv})$ is employed to update the value function and policy, which can be expressed as follows:

$$\begin{aligned}A_i(a_i(t), o(t); \theta, \theta_v) &= R_i(t) - V_i(o(t); \theta_{iv})\\ &= \sum_{m=0}^{k-1}\gamma^m r_i(t+i) + \gamma^k V_i(o(t+k); \theta_{iv})\\ &\quad - V_i(o(t); \theta_{iv})\end{aligned} \tag{19}$$

After each update of the global network, its CNN parameters will be assigned to the agents of each thread to ensure that each thread shares a common policy.

In the training process, the policy loss function $f_i^{\pi_i}(\theta_i)$ and estimated value loss function $f_i^v(\theta_i)$ are formulated as follows:

$$\begin{aligned}f_i^{\pi_i}(\theta) &= \log\pi_i(a_i(t)|o(t);\theta_i)(R_i(t) - V_i(o(t);\theta_{iv}))\\ &\quad + \beta H(\pi_i(o(t);\theta_i))\end{aligned} \tag{20}$$

$$f_i^v(\theta_{iv}) = (R_i(t) - V_i(o(t);\theta_{iv}))^2 \tag{21}$$

where $H(\pi_i(o(t);\theta))$ is the entropy term and β is the weight parameter of entropy regularization. Adding the entropy term in (20) could promote exploration in the training process. The entropy regularization term and advantage function are combined to form the optimization goal of the policy output, which is minimizing the policy loss function. The minimizing value loss function (21) is set as the optimization goal of the value output, where the value loss function is derived from the estimated advantage function.

Algorithm 1: A3C Based MA-DRL Algorithm for SFC and SoC Balancing Issues With ET Mechanism in BESSs

foreach *agent i* **do**
  Initialization :
    Global-shared counter $T_i = 0$
    Global-shared CNN parameter θ_i and θ_{iv}
    Thread-characteristic CNN parameter θ_i' and θ_{iv}'
  end
end
foreach *episode* **do**
  foreach *thread of agent i* **do**
    while $T_i \le T_{max}$ **do**
      Reset gradients: $d\theta_i \leftarrow 0$ and $d\theta_{iv} \leftarrow 0$.
      Reset thread step counter $t \leftarrow 0$, $t_{start} \leftarrow t$
      update parameters $\theta_i' = \theta_i$ and $\theta_{iv}' = \theta_v$ synchronously.
      while $t - t_{state} < t_{max}$ *or* $s_i(t)$ *is not terminal* **do**
        Update event-triggered mechanism according to global observation $o(t)$
        Get $s_i(t)$ based on event-triggered condition (14) and the observation from other cooperative agents
        Carry out $a_i(t)$ based on the policy network $\pi_i(a_i(t)|o(t);\theta_i')$
        Obtain reward immediate $r_i(t)$ according to (16) and new observation $o(t+1)$
        $t \leftarrow t+1, T \leftarrow T+1$
      end
      Calculate R_i of the last global observation $o(t)$
      $R_i = \begin{cases} 0 & \text{at terminal} \\ V_i\big(o(t),\theta_{iv}'\big) & \text{at nonterminal}\end{cases}$
      foreach $i \in t_1, \ldots, t_{start}$ **do**
        $R_i \leftarrow r_i + \lambda R_i$
        $d\theta_i \leftarrow d\theta_i + \nabla_{\theta_i'}\log\pi_i\big(a_i|o_i;\theta_i'\big)(R_i - V_i(o_i;\theta_{iv}')) + c\nabla_{\theta_i'}H(\pi_i(o_i;\theta_i))$
        $d\theta_{iv} \leftarrow d\theta_{iv} + \nabla_{\theta_{iv}}(R_i - V_i(o_i;\theta_{iv}'))^2/2$
      end
      Asynchronous updating of θ_i using $d\theta_i$ and of θ_{iv} using $d\theta_{iv}$
    end
  end
end

TABLE I
PARAMETER OF FOUR-BUS BESSS

BESS (Bus No.)	1	2	3	4
η_i	0.981	0.982	0.99	0.993
Q_i	120	118	110	125
V_i	220	222	221	220
K_i^w	0.05	0.07	0.08	0.06

IV. SIMULATIONS

In this section, a four-bus BESS system is taken as the simulation example to investigate the proposed A3C based MA-DRL algorithm and SINR-based event-triggered mechanism for the SFC and SoC balancing coordination subject to DoS attacks. The parameters of BESSs are presented in Table I as follows.

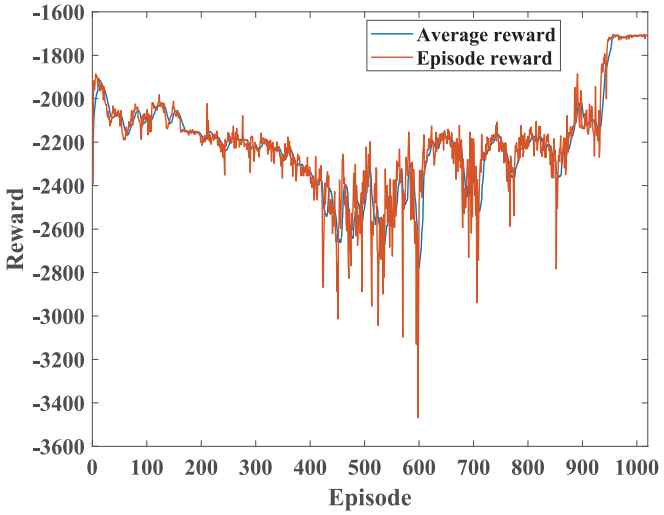

Fig. 3. Global episode reward and average reward for achieving SFC and SoC balancing issues.

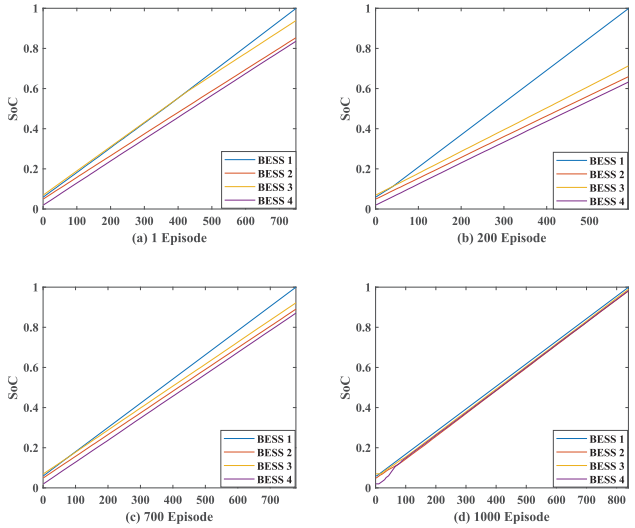

Fig. 4. SoC charging process for four-bus BESSs.

As there are different BESSs, e.g., new energy vehicles of various brands, connected to each bus in reality, the charge/discharge capacity parameters Q_i of their own are different. It is necessary to enable different batteries to complete the charging/discharging process in a relatively concentrated time, which is conducive to peak shaving/shifting of the power system and extending battery life while ensuring the overall active power requirements. Otherwise, if there is a battery that completes the charging/discharging process ahead of time, it will be cut out of the power system, and the remaining batteries will not be able to meet the overall active power requirement. The feasible range of SoC value for four BESSs is $(0, 1)$, where 0 and 1 mean that the battery is exhausted or fully charged. For learning algorithm, we select $\rho = 0.001$, $\gamma = 0.99$, $\beta = 32$ and $K_{i1} = K_{i2} = -2, K_{i3} = K_{i4} = -1, i \in \{1, 2, 3, 4\}$ for each agents. $J + (\sigma_i)^2 = 10$ in SINR function and

$$T_i = \begin{cases} 10 & \text{data releases at } i \text{ BESS} \\ 0 & \text{no data releases at } i \text{ BESS} \end{cases} \tag{22}$$

$f(\text{SINR}_i(t)) = \epsilon_n e^{-\text{SINR}_i(t)}$, $\epsilon_n = 5$, which means the communication channels are fast-fading [35]. The initial dynamic threshold of event-triggered condition (14) is chosen as $\sigma_{im} = 0.0001$.

It is assumed that the BESSs are running in the charging mode and the requirement for overall active power is -200kw. The discharging process is the reverse process of charging, and its physical process is similar, so this paper will not repeat the simulation of its process. We assume that initial SoC values of four-bus BESSs are $\{0.04, 0.05, 0.03, 0.02\}$ and the initial frequency values (p.u.) are $\{1.3, 1.1, 0.8, 0.3\}$. We set the maximum episode as 1100 and the maximum step of each episode as 1000. The parallel thread of each agent is 4. The termination condition of each training episode is that the maximum number of training steps is reached or any battery is fully charged. The global episode reward and average reward of the A3C based MA-DRL algorithm for achieving SFC and SoC balancing subject to DoS attacks are displayed in Fig. 3.

Since the cooperative Markov games is considered in this paper, it is necessary to analyze the variation of global rewards for all agents in the training process. It is shown in Fig. 3 that the episode reward is trained from -2400. After exploration and training, the average reward experienced ups and downs, and some bad policies are discarded in this process. At the end of the training, the average reward converges near -1700, which means that the BESSs have obtained a relatively optimal policy. We apply the A3C based MA-DRL algorithm to deal with the SFC and SoC balancing issues simultaneously. In order to show the training process and effect more clearly, we take the $\{1th, 200th, 700th, 1000th\}$ episodes as examples to show the charging process as shown in Fig. 4.

Since the charging capacity of each battery is different, the SoC of each BESSs cannot converge according to its charging characteristics during the initial training episode as shown in Fig. 4(a). During the training process, the agent may explore a relatively poor policy, resulting in a decrease in the cumulative reward function and deterioration of charging performance, which cannot achieve consensus of SoC value in Fig. 4(b). This policy will be abandoned. After continuous exploration and training, the cooperative multi-agent gradually mastered enough experience to apply different powers to different batteries to achieve SoC balancing issue. Until the end of training as shown in Fig. 4(d), even if the initial SoC and charging capabilities are different, the SoC values of all BESSs can converge, which means that all batteries can be fully charged at the same time while meeting the requirement of overall active power. Since the agent has to apply different active power to the corresponding SoC value, the variation of active power will cause frequency fluctuations due to primary frequency control. Thus, it is necessary to consider SFC and SoC balancing issues simultaneously. Fig. 5 shows the SFC training process while training the SoC balancing issue.

The goal of SFC training is to make the frequency stable as fast as possible while completing the SoC training goal based on reward function (16). If there is no training of SFC issue, the frequency fluctuation due to the variation of active power is shown in Fig. 5(a), which will cause deterioration in

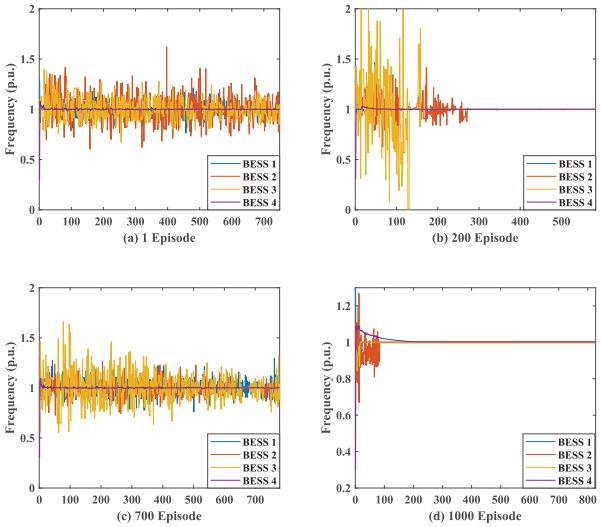

Fig. 5. SFC process for four-bus BESSs.

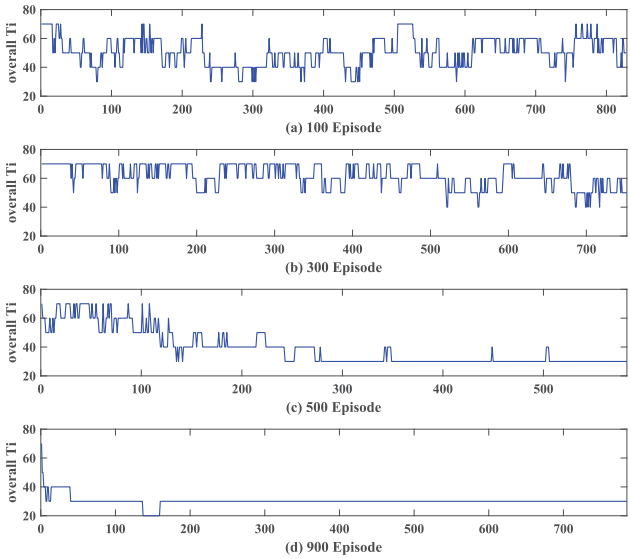

Fig. 6. Overall transmission power of four-bus BESSs.

power quality and affect the synchronization of motor speeds. With the training process, the agent may explore a policy that reduces the frequency convergence time but the fluctuation range is greater as shown in Fig. 5(b). This policy of reducing the convergence time by expanding the oscillation amplitude is not optimal, and it will reduce the cumulative reward in Fig. 3. Obviously, it would be abandoned. As the episode of later training increases, the oscillation amplitude of the frequency becomes smaller in Fig. 5(c). In Fig. 5(d), it is presented that the frequency deviation caused by the change of active power can be recovered in a relatively short time, which means the A3C based MA-DRL algorithm has obtained a relatively optimal policy to realize the SFC and SoC balancing issues subject to DoS attacks at the same time. In this paper, we propose an SINR-based event-triggered mechanism to cope with DoS attacks. The proposed event-triggered strategy dynamically adapts the release frequency according to SINR, while ensuring system performance, in order to achieve the purpose of reducing the occupied network bandwidth and further reducing PER induced by DoS attacks. According to the definition of transmission power in (22), taking the 100, 300, 500, 900th episode as examples, the overall transmission power of eight channels in four buses is shown in Fig. 6.

Based on (22), when the communication channel sends data, the transmission power needs to be applied to facilitate the delivery. Therefore, reducing the amount of transmission packets in each BESS can reduce the overall transmission power and improve the network congestion environment, which can further reduce the threat of DoS attacks. In Fig. 6, it can be seen that as the training episode increases, the overall transmission power decreases and gradually stabilizes, which reduces the waste of communication resources while ensuring the realization of SFC and SoC balancing issues. SINR can be calculated according to the overall communication power, and the probability of transmission being DoS attacked can be further derived as shown in Fig. 7.

As the episode of training increases, the probability of transmission being DoS attacked decreases, reflecting the

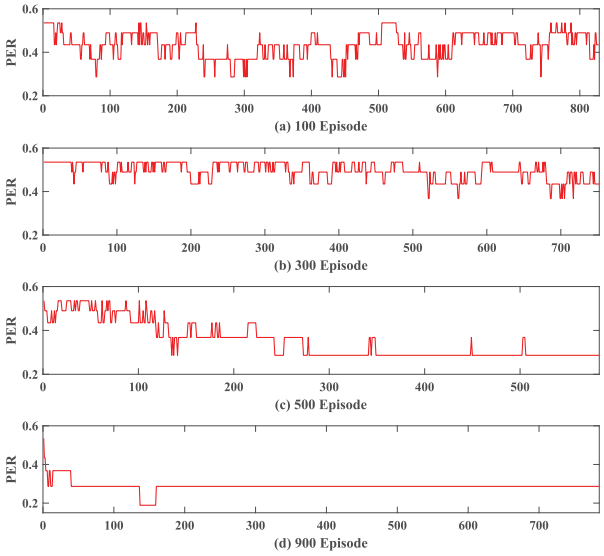

Fig. 7. Overall PER induced by DoS attacks.

superiority of the proposed event-triggered mechanism for scheduling communication resources and reducing the impact of DoS attacks as shown in Fig. 7. In order to show the superiority of the dynamic event-triggered mechanism proposed in this paper more directly, we take the SFC in the BESS 1 as an example to show the release instant and interval as shown in Fig. 8.

It is illustrated in Fig. 8 that as the episode of training increases, the number of releases is gradually reduced to prevent the waste of communication resources. At the 900th training episode, the data release is only in the initial control stage, which means that the training policy can not only quickly achieve SFC but also reduce unnecessary control instants.

The active power sharing results in four-bus BESSs are displayed in Fig. 9. In Fig. 9, it is observed that the active power sharing process is accomplished after 1000 episodes in

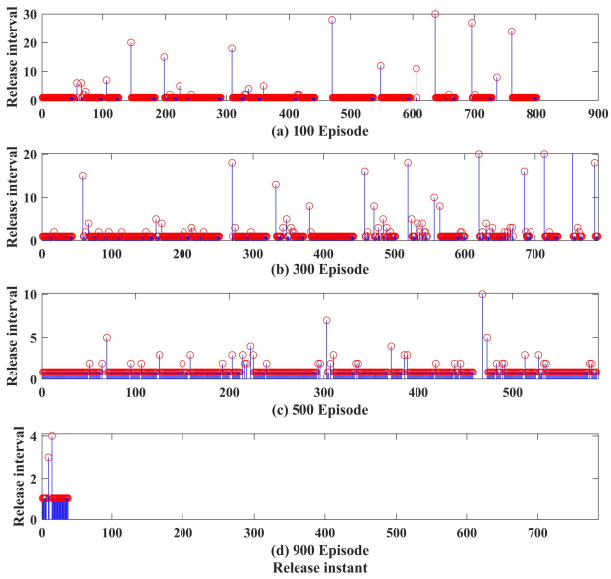

Fig. 8. Release instant and interval of the BESS 1.

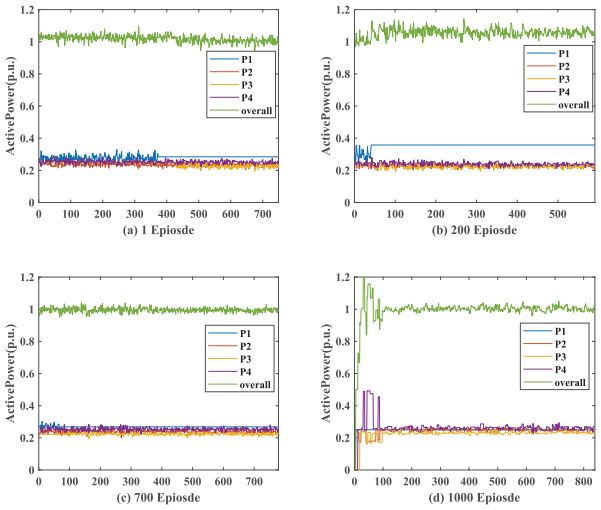

Fig. 9. Active power sharing process for four-bus BESSs.

training. The BESS 4 has the lower initial SoC value as shown in Fig. 4 and P_4 is higher compared to the active power of the rest BESSs in the beginning stage so that the SoC of BESS 4 could reach to an average value of all the four BESSs. The average steady-state ratios of the SoC of the four BESSs are respectively {1:0.989:0.920:1.041}. In view of the parameter setting in Table I and steady-state power sharing (10), the ratio is reasonable. Besides, the limitation of overall active power in (9) is achieved in the stable charging stage as shown in Fig. 9. In order to reflect the superiority of the event-triggered mechanism proposed in this paper, the average amount of released data per training episode under different triggered mechanisms is shown in Table II.

The σ_{im} means the constant threshold in [33], the initial threshold in [26] and this work. According to the data in Table II, it can be clearly obtained that if the threshold is chosen as 0, the event-triggered mechanisms [26], [33] and (14) degenerate to the periodic time-triggered mode. As

TABLE II
AVERAGE AMOUNT OF RELEASED DATA BASED ON DIFFERENT TRIGGERED SCHEME

σ_{im}	0	0.0001	0.001	0.005
Time-Triggered [5]	737	–	–	–
constant threshold-based ET [33]	737	635	589	518
Adaptive ET [26]	737	573	526	433
this work	737	450	436	268

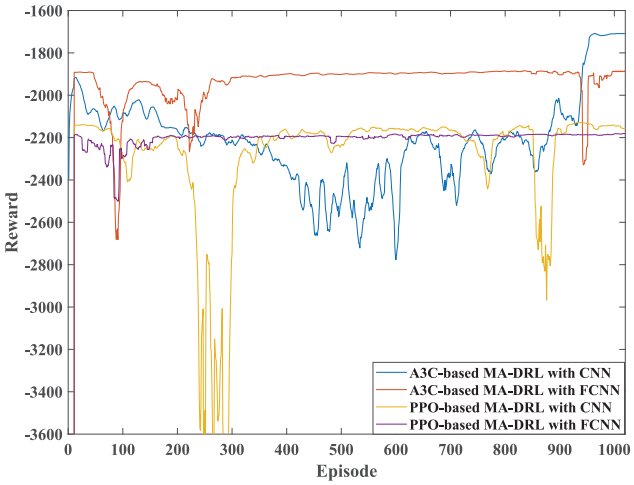

Fig. 10. Comparison of training performance between four methods.

the threshold increases, the average amount of released data per training episode gradually decreases. Because the event-triggered mechanism is more proactive in dealing with DoS attacks, it can release fewer times corresponding to each threshold, which could increase the average release interval and save more communication resources.

Proximal policy optimization (PPO) is a novel policy gradient algorithm, which tries to obtain a new policy in each iteration to minimize the loss function in the training process, and to ensure that each newly calculated policy is not much different from the original one based on the actor-critic framework. We have compared the training performance between the A3C-based MA-DRL algorithm designed in this paper and the PPO-based MA-DRL algorithm in [40] with fully connected neural network (FCNN) and convolutional neural networks (CNN) in Fig. 10. The steady-state average accumulated reward and training time with 1000 episodes of four methods are listed in Table III. It is shown in Fig. 10 that after 1000 episodes, the average accumulated reward of four methods can converge, where the MA-DRL algorithm with FCNN has a faster convergence rate than the corresponding one with CNN, but the steady-state reward is relatively lower. This is because FCNN has a simpler structure than CNN, which makes it easier to be trained but being more difficult to accurately acquire model features and obtain the optimal policy. In Table III, the MA-DRL algorithm with CNN has a higher average accumulated reward than the corresponding one with FCNN, and the A3C-based MA-DRL with CNN has the largest average accumulated reward, which means the algorithm proposed in this paper can obtain a comparatively optimal policy through training. With respect

TABLE III
REWARD AND TRAINING TIME WITH 1000 EPISODES OF FOUR METHODS

Method	Reward	Time(Second)
A3C-based MA-DRL with CNN	-1710.03	7512
A3C-based MA-DRL with FCNN	-1885.14	6265
PPO-based MA-DRL with CNN	-2142.17	8132
PPO-based MA-DRL with FCNN	-2186.21	6819

to training time, the A3C-based MA-DRL algorithm has a shorter training time than the PPO-based MA-DRL algorithm when using the same neural network, which means the method proposed in this paper is more efficient. In summary, the A3C-based MA-DRL algorithm with CNN obtains better training performance and policy than the PPO-based MA-DRL algorithm while it sacrifices marginally training time compared to FCNN. Considering the tradeoff between training time and performance, the algorithm designed in this paper is comparatively optimal.

V. CONCLUSION

In this paper, we propose a novel A3C based MA-DRL algorithm to cope with the SFC and SoC balancing problems simultaneously with the centralized learning and decentralized execution framework subject to DoS attacks in the local communication network between the secondary controller and remote sensors. In this control scheme, the CNN is applied to estimate the value function and policy function in each cooperative agent with the asynchronous mode. Besides, in order to deal with the DoS attacks proactively, an SINR-based dynamic event-triggered communication strategy is designed to mitigate the influence of the DoS attacks. Finally, simulation results on a four-bus BESS system verify the proposed A3C based MA-DRL algorithm can achieve the SFC and SoC balancing simultaneously for multiple heterogeneous BESSs. Meanwhile, it is demonstrated that compared with other event-triggered mechanisms, the proposed SINR-based approach significantly reduces the amount of released data while ensuring system performance and the occupation of communication bandwidth, correspondingly to achieve the goal of alleviating the channel congestion caused by DoS attacks.

REFERENCES

[1] M. T. Lawder *et al.*, "Battery energy storage system (BESS) and battery management system (BMS) for grid-scale applications," *Proc. IEEE*, vol. 102, no. 6, pp. 1014–1030, Jun. 2014.

[2] X. Lu, K. Sun, J. M. Guerrero, J. C. Vasquez, and L. Huang, "State-of-charge balance using adaptive droop control for distributed energy storage systems in DC microgrid applications," *IEEE Trans. Ind. Electron.*, vol. 61, no. 6, pp. 2804–2815, Jun. 2014.

[3] L. Xing *et al.*, "Dual-consensus-based distributed frequency control for multiple energy storage systems," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6396–6403, Nov. 2019.

[4] M. Farrokhabadi *et al.*, "Microgrid stability definitions, analysis, and examples," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 13–29, Jan. 2020.

[5] C. Deng, Y. Wang, C. Wen, Y. Xu, and P. Lin, "Distributed resilient control for energy storage systems in cyber-physical microgrids," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 1331–1341, Feb. 2021.

[6] J. W. Simpson-Porco, Q. Shafiee, F. Dörfler, J. C. Vasquez, J. M. Guerrero, and F. Bullo, "Secondary frequency and voltage control of islanded microgrids via distributed averaging," *IEEE Trans. Ind. Electron.*, vol. 62, no. 11, pp. 7025–7038, Nov. 2015.

[7] J. Pahasa and I. Ngamroo, "PHEVs bidirectional charging/discharging and SoC control for microgrid frequency stabilization using multiple MPC," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 526–533, Mar. 2015.

[8] Y. Xu, M. Fang, Z.-G. Wu, Y.-J. Pan, M. Chadli, and T. Huang, "Input-based event-triggering consensus of multiagent systems under denial-of-service attacks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 4, pp. 1455–1464, Apr. 2020.

[9] S. Lakshminarayana, A. Kammoun, M. Debbah, and H. V. Poor, "Data-driven false data injection attacks against power grids: A random matrix approach," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 635–646, Jan. 2021.

[10] X. Li and K. W. Hedman, "Enhancing power system cyber-security with systematic two-stage detection strategy," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1549–1561, Mar. 2020.

[11] L. Ding, Q.-L. Han, B. Ning, and D. Yue, "Distributed resilient finite-time secondary control for heterogeneous battery energy storage systems under denial-of-service attacks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4909–4919, Jul. 2020.

[12] Y.-J. Kim and J. Wang, "Power hardware-in-the-loop simulation study on frequency regulation through direct load control of thermal and electrical energy storage resources," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2786–2796, Jul. 2018.

[13] H. Liu and W. Wu, "Online multi-agent reinforcement learning for decentralized inverter-based volt-VAR control," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 2980–2990, Jul. 2021.

[14] Y. Xu and W. Liu, "Novel multiagent based load restoration algorithm for microgrids," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 152–161, Mar. 2011.

[15] D. Cao, J. Zhao, W. Hu, F. Ding, Q. Huang, and Z. Chen, "Attention enabled multi-agent DRL for decentralized volt-VAR control of active distribution system using PV inverters and SVCs," *IEEE Trans. Sustain. Energy*, vol. 12, no. 3, pp. 1582–1592, Jul. 2021.

[16] M. Shin, D.-H. Choi, and J. Kim, "Cooperative management for PV/ESS-enabled electric vehicle charging stations: A multiagent deep reinforcement learning approach," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3493–3503, May 2020.

[17] D. Cao *et al.*, "Data-driven multi-agent deep reinforcement learning for distribution system decentralized voltage control with high penetration of PVs," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 4137–4150, Sep. 2021, doi: 10.1109/TSG.2021.3072251.

[18] S. Wang *et al.*, "A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4644–4654, Nov. 2020.

[19] M. Kamruzzaman, J. Duan, D. Shi, and M. Benidris, "A deep reinforcement learning-based multi-agent framework to enhance power system resilience using shunt resources," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5525–5536, Nov. 2021, doi: 10.1109/TPWRS.2021.3078446.

[20] D. Yue, E. Tian, and Q.-L. Han, "A delay system method for designing event-triggered controllers of networked control systems," *IEEE Trans. Autom. Control*, vol. 58, no. 2, pp. 475–481, Feb. 2013.

[21] Y. Tang, D. Zhang, P. Shi, W. Zhang, and F. Qian, "Event-based formation control for nonlinear multiagent systems under DoS attacks," *IEEE Trans. Autom. Control*, vol. 66, no. 1, pp. 452–459, Jan. 2021.

[22] X. Su, X. Liu, and Y.-D. Song, "Event-triggered sliding-mode control for multi-area power systems," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6732–6741, Aug. 2017.

[23] K.-D. Lu, G.-Q. Zeng, X. Luo, J. Weng, Y. Zhang, and M. Li, "An adaptive resilient load frequency controller for smart grids with DoS attacks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 4689–4699, May 2020.

[24] Y. Wang, C. Deng, D. Liu, Y. Xu, and J. Dai, "Unified real power sharing of generator and storage in islanded microgrid via distributed dynamic event-triggered control," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 1713–1724, May 2021.

[25] J. Lai, X. Lu, X. Yu, and A. Monti, "Stochastic distributed secondary control for AC microgrids via event-triggered communication," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 2746–2759, Jul. 2020.

[26] C. Peng, J. Zhang, and H. Yan, "Adaptive event-triggering H_∞ load frequency control for network-based power systems," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1685–1694, Feb. 2018.

[27] N. G. B. Amma, S. Selvakumar, and R. L. Velusamy, "A statistical approach for detection of denial of service attacks in computer networks," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2511–2522, Dec. 2020.

[28] S. Hu, D. Yue, X. Chen, Z. Cheng, and X. Xie, "Resilient H_∞ filtering for event-triggered networked systems under nonperiodic dos jamming attacks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 3, pp. 1392–1403, Mar. 2021.

[29] S. Hu, D. Yue, X. Xie, X. Chen, and X. Yin, "Resilient event-triggered controller synthesis of networked control systems under periodic DoS jamming attacks," *IEEE Trans. Cybern.*, vol. 49, no. 12, pp. 4271–4281, Dec. 2019.
[30] J. M. Guerrero, J. C. Vasquez, J. Matas, L. G. de Vicuna, and M. Castilla, "Hierarchical control of droop-controlled AC and DC microgrids—A general approach toward standardization," *IEEE Trans. Ind. Electron.*, vol. 58, no. 1, pp. 158–172, Jan. 2011.
[31] Q. Shafiee, J. M. Guerrero, and J. C. Vasquez, "Distributed secondary control for islanded microgrids—A novel approach," *IEEE Trans. Power Electron.*, vol. 29, no. 2, pp. 1018–1031, Feb. 2014.
[32] G. Chen and Z. Guo, "Distributed secondary and optimal active power sharing control for islanded microgrids with communication delays," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2002–2014, Mar. 2019.
[33] L. Xing *et al.*, "Distributed state-of-charge balance control with event-triggered signal transmissions for multiple energy storage systems in smart grid," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 8, pp. 1601–1611, Aug. 2019.
[34] C. Peng, J. Li, and M. Fei, "Resilient event-triggering H_∞ load frequency control for multi-area power systems with energy-limited dos attacks," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 4110–4118, Sep. 2017.
[35] H. Yuan, Y. Xia, and H. Yang, "Resilient state estimation of cyber-physical system with multichannel transmission under DoS attack," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 11, pp. 6926–6937, 2021, doi: 10.1109/TSMC.2020.2964586.
[36] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York, NY, USA: McGraw-Hill, 2007.
[37] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Proc. 11th Int. Conf. Mach. Learn. (ICML-94)*, Jul. 1994, pp. 157–163.
[38] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," in *Advances in Neural Information Processing Systems*, vol. 30. Red Hook, NY, USA: Curran Assoc., Inc., 2017, pp. 6379–6390.
[39] M. S. Munir, S. F. Abedin, N. H. Tran, Z. Han, E.-N. Huh, and C. S. Hong, "Risk-aware energy scheduling for edge computing with microgrid: A multi-agent deep reinforcement learning approach," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 3, pp. 3476–3497, Sep. 2021, doi: 10.1109/TNSM.2021.3049381.
[40] D. Guo, L. Tang, X. Zhang, and Y.-C. Liang, "Joint optimization of handover control and power allocation based on multi-agent deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13124–13138, Nov. 2020.

Pengcheng Chen received the B.E. degree in electrical engineering and its automation and the M.E. degree in control theory and engineering from the Zhejiang University of Technology, Hangzhou, China, in 2018 and 2021, respectively. He is currently pursuing the Ph.D. degree with the Department of Electronics, Carleton University, Ottawa, ON, Canada.

His current research interests include reinforcement learning, secondary control of microgrids, and adaptive event-triggered strategy.

Shichao Liu (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in control engineering from Harbin Engineering University, Harbin, China, in 2007 and 2010, respectively, and the Ph.D. degree in electrical and computer engineering from Carleton University, Ottawa, ON, Canada, in 2014.

He is currently an Assistant Professor with the Department of Electronics, Carleton University. His research interests include machine learning and reinforcement learning for cyber–physical energy systems with applications in microgrids and large-scale power systems. He is an Associate Editor of IEEE ACCESS and a member of the editorial board of Smart Cities.

Bo Chen (Member, IEEE) received the B.S. degree in information and computing science from the Jiangxi University of Science and Technology, Ganzhou, China, in 2008, and the Ph.D. degree in control theory and control engineering from the Zhejiang University of Technology, Hangzhou, China, in 2014.

He joined the Department of Automation, Zhejiang University of Technology in 2018, where he is currently a Professor. He was a Research Fellow of the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from 2014 to 2015 and from 2017 to 2018. He was also a Postdoctoral Research Fellow of the Department of Mathematics, City University of Hong Kong, Hong Kong, from 2015 to 2017. His current research interests include information fusion, distributed estimation and control, cyber–physical systems security, and networked fusion systems.

Dr. Chen was a recipient of the Outstanding Thesis Award of Chinese Association of Automation in 2015.

Li Yu (Member, IEEE) received the B.S. degree in control theory from Nankai University, Tianjin, China, in 1982, and the M.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 1988 and 1999, respectively.

He is currently a Professor with the College of Information Engineering, Zhejiang University of Technology. He has successively presided over 20 research projects. He has authored or coauthored five academic monographs, one textbook, and over 300 journal articles. He has also been authorized over 100 patents for invention and granted five scientific and technological awards. His current research interests include robust control, networked control systems, cyber–physical systems security, and information fusion.