Destabilizing Attack and Robust Defense for Inverter-Based Microgrids by Adversarial Deep Reinforcement Learning

Yu Wang, *Member, IEEE* and Bikash Pal, *Fellow, IEEE*

***Abstract*—The droop controllers of inverter-based resources (IBRs) can be adjustable by grid operators to facilitate regulation services. Considering the increasing integration of IBRs at power distribution level systems like microgrids, cyber security is becoming a major concern. This paper investigates the data-driven destabilizing attack and robust defense strategy based on adversarial deep reinforcement learning for inverter-based microgrids. Firstly, the full-order high-fidelity model and reduced-order small-signal model of typical inverter-based microgrids are recapitulated. Then the destabilizing attack on the droop control gains is analyzed, which reveals its impact on system small-signal stability. Finally, the attack and defense problems are formulated as Markov decision process (MDP) and adversarial MDP (AMDP). The problems are solved by twin delayed deep deterministic policy gradient (TD3) algorithm to find the least effort attack path of the system and obtain the corresponding robust defense strategy. The simulation studies are conducted in an inverter-based microgrid system with 4 IBRs and IEEE 123-bus system with 10 IBRs to evaluate the proposed method.**

***Index Terms*—Destabilizing attack, microgrids, inverter-based resources, deep reinforcement learning, adversarial training.**

I. Introduction

The power system is facing the uphill challenge of high-level penetration of renewable generation, in order to meet the net-zero carbon target in the energy sector [1], [2]. In distribution-level microgrid systems, a large-scale of inverter-based resources (IBRs) is being connected to the power network in a distributed way. Different from bulk power systems dominated by synchronous generators, the dynamics of these inverter-based systems are determined by the control modes of power electronic interfaces. Besides, the power electronic devices present a much faster response than synchronous generators, which means the time scale of the network dynamics is comparable and cannot be ignored in stability analysis. To facilitate the regulation services in a time-varying system environment, certain control parameters of IBRs become adjustable or dispatchable. Like a double-edged sword, the flexibility brought by the user-defined control systems of power converters will also increase the attack surface. Therefore, the vulnerability and cyber-security for inverter-integrated power systems are emerging but important problems to be investigated.

The cyber-security of bulk power systems with multi-machines has raised concerns for a long time. Cyber-security of different processes in power systems has been studied, such as state estimation [3], power dispatch [4], and automatic generation control [5]. In addition, a few works consider cyber-attacks for destabilizing dynamic power systems. In the early stage, the author in [6] introduces the *destabilizing attack* of power systems through the state-feedback controller. The synchronous generators are divided into *control group* manipulated by malicious attackers, and *target group* to be destabilized. The attack aims to shift certain sensitive eigenvalues from the left into the right plane. This method is later applied to mixed-source microgrids [7]. In recent works, dynamic load-altering attacks are studied, as the wide adoption of demand response schemes increases the attack surface. In [8], the attack on the dynamic loads aims to destabilize the power systems, where the victim loads are changed based on the feedback of system frequency. A non-convex optimization problem is formulated to determine the minimum amount of load to be protected at each bus. In [9], the latency attack on the automatic generation control of the power system and its impact on system stability is studied. A parameter tuning method based on an exhaustive and heuristic search is proposed to maximize the stability region under such attack.

In the meantime, the cyber-security problem has also raised much attention in power electronics-enriched systems like microgrids. The wide integration of IBRs increases system flexibility while decreasing system security. A large amount of work has been conducted on the secondary control systems for microgrids, as its attack surface is enlarged with the utilization of the communication systems [10]. The impact of typical attacks such as false data injection (FDI) and denial-of-service (DoS) are investigated [11], [12]. Methodologies have been provided for cyber attack prevention, detection, isolation, and mitigation for network-controlled microgrids. The resilient control and detection indexes are designed considering the specific consensus algorithms in the secondary control of microgrids. It is noted that the FDI and DoS attacks mainly influence the system operational points targeting to make the system violate the security boundary [13].

It can be found that prior works on destabilizing attacks focused on synchronous generator dominated power systems [6], [8], [9]. There is minimal work on the cyber-attacks target

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. MSCA-IF-2020, 101026657.

Y. Wang and B. Pal are with the Control and Power Group, Department of Electrical and Electronic Engineering, Imperial College London, UK, 639798. (e-mail: yu.wang@imperial.ac.uk, b.pal@imperial.ac.uk).

to small-signal stability and its defense mechanism of inverter-based systems like microgrids. It motivates us to further study this problem. Specifically, the least effort attack with minimal droop parameter change to destabilize the inverter-based microgrids is studied. It will help to understand the system's vulnerable parameters and manifestation under destabilizing attacks. In addition, it also contributes to developing corresponding defensive mechanisms to mitigate its impact.

These kinds of attack and defense problems can be formulated as dynamic programming or optimal control problems. However, these problems usually involve non-linear dynamics of system models and non-convexity in solving system eigenvalues. It innovates us to apply the data-driven method based on deep reinforcement learning (DRL) to find online approximate solutions to such problems. The DRL approaches have been widely used for power engineering problems, such as voltage control [14], frequency control [15], energy management [16], etc. A literature review of DRL application in power systems is provided in [17]. By interacting with the dynamic environment, the DRL algorithms can train the deep neural networks (DNNs) based agents to find an optimal control policy. Based on the policy, DRL methods can be divided into deterministic policy, e.g. deep deterministic policy gradient (DDPG), and stochastic policy, e.g. proximal policy optimization (PPO) and soft actor-critic (SAC). The candidates of DRL have been applied to address cyber-security problems of microgrids and power systems in some recent works [18]–[20]. In [18], a multi-agent deep Q network approach is proposed to detect the vulnerable spots in the index-based detection schemes for the secondary control in islanded DC microgrids. In [19], DRL based method is proposed for providing optimal defense strategy for microgrids subject to FDI on the load demand. In [20], an asynchronous advantage actor-critic (A3C) based multi-agent DRL is proposed to provide resilient control for the secondary control of microgrids to alleviate the impact of DoS attacks. In addition, the method of adversarial reinforcement learning has been proposed to find robust control solutions for voltage var control problems in power distribution networks with uncertainty in the environment [21]. The adversarial training of DRL agents has been proposed for robust continuous control with attackers in cyber-physical power systems [22]. This approach demonstrates its potential for addressing the destabilizing attack and robust defense problem in inverter-based systems.

In this paper, the cyber-attack and defense strategy in inverter-based microgrids is studied systematically. Specifically, the impacts of destabilizing attacks on droop control gains to the system stability are analyzed. The attack functions to shift the system shrinking the small-signal stability region by manipulating droop gains. The analysis reveals that such attacks can be defended by changing sensitive droop gains of the system. Then the least effort attack (LEA) and its defense problems are introduced correspondingly. The attack and defense problems are formulated as Markov decision process (MDP) and adversarial MDP (AMDP). The twin delayed deep deterministic policy gradient (TD3), as a deterministic policy DRL method, is proposed to identify the dynamic LEA for inverter-based systems. Compared to stochastic policy, the agent with deterministic policy by TD3 can provide a deterministic action to adjust the droop gains in the dynamic system. Besides, an adversarial reinforcement learning framework is adopted to find the dynamic and robust defense strategy under LEA. The distinct contributions of this paper compared to existing works are:

- Considering the small-signal stability in inverter-based systems, the destabilizing attack is modelled and analyzed for the first time.
- The attack and defense problems are formulated as finding the optimal combination of droop gains within attack and defense sets in inverter-based systems.
- The TD3 algorithm is adopted for training the attack agents, while the robust defense strategy is generated by adversarial training between attack and defense agents.

II. System Modelling

To investigate the destabilizing attack on the system stability, the dynamic model of multi-inverter microgrid systems is presented. Based on the full-order high-fidelity model, the reduced-order small-signal model can be derived [23]–[25]. They are used to calculate the system trajectory as well as the trace of eigenvalues under cyber-attack. The system model consists of the dynamics of inverters, network and loads, and the transformation between local and common frames. The network dynamics are taken into account in the system as the IBRs respond quite fast as compared to synchronous generators.

Notation: $\mathbf{1}_N$ stands for N-dimensional identity matrix. $\mathbf{0}_N$ stands for N-dimensional zero matrix. For vector $x \in \mathbb{R}^N$, $x = col\{x_1, \ldots, x_N\}$. $\mathcal{X} = diag\{x_1, x_2, ..., x_N\}$ denotes a N-dimensional diagonal matrix with x_1, $\ldots$, x_N on its diagonal elements.

A. Modelling of Inverter-based Microgrids

A typical inverter-based microgrid with multiple IBRs governed by grid-forming and droop control is considered in

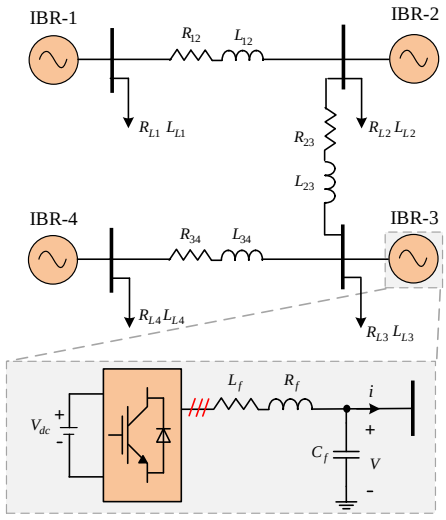

Fig. 1. A inverter-based microgrid with 4 IBRs and its power electronic interface.

this paper, as demonstrated in Fig. 1. The power network of the system can be represented by a complex-weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the nodes $\mathcal{V}$ represent the buses, and the edges $\mathcal{E}$ represent the line connections. The loads and inverters are connected sparsely at each bus.

For generality, it is considered that the DC-side voltage is well maintained at the primary side. The inverter dynamics can be modelled with a continuous average model as its high switching frequency. The modelling is conducted in *dq* frame which can be converted to *abc* frame by Park Transformation. The local *dq* frame can be transferred into a common reference *DQ* frame as follows [23]:

$$x_{DQ}(t) = \mathcal{T}(\delta(t)) x_{dq}(t) \tag{1}$$

where $x_{DQ} = [x_D, x_Q]^T$, $x_{dq} = [x_d, x_q]^T$, $\mathcal{T}(\delta(t)) = \begin{bmatrix} \cos\delta(t) & -\sin\delta(t) \\ \sin\delta(t) & \cos\delta(t) \end{bmatrix}$.

The angle of ith inverter is calculated by:

$$\dot{\delta}_i(t) = \omega_i(t) - \omega_{com}(t) \tag{2}$$

where ω_{com} is the common reference frequency.

1) *Inverter Modelling:*. The droop control for power inverter of IBRs is designed with the philosophy of emulating the behavior of synchronous generators to share the load demand based on frequency deviation. Similarly, the reactive power can be shared by droop control with the voltage magnitude. Considering a first-order filter in the power calculation process, they can be represented as [25]:

$$\tau\dot{\omega}_i = -\omega_i + \omega_n - m_i P_i \tag{3}$$

$$\tau\dot{V}_i = -V_i + V_n - n_i Q_i \tag{4}$$

where ω_i, V_i are frequency and voltage references for inner control loops, ω_n, V_n are nominal value of frequency and voltage, P_i, Q_i are measured real and reactive power, m_i, n_i are corresponding droop gains. $\tau = \frac{1}{\omega_c}$ is the low-pass filter time constant for the power measurement, ω_c is the cut-off frequency. It is noted that the output voltage magnitude is aligned to the local d-axis of the inverter reference frame ($V_i = v_{di}^*$), while the q-axis reference is zero ($v_{qi}^* = 0$). The droop gains are typically selected based on allowable frequency and voltage range, as follows [23]:

$$m_i \leq \frac{\overline{\omega_i} - \underline{\omega_i}}{\overline{P_i} - \underline{P_i}} \tag{5}$$

$$n_i \leq \frac{\overline{V_i} - \underline{V_i}}{\overline{Q_i} - \underline{Q_i}} \tag{6}$$

where $\overline{\omega_i}$, $\underline{\omega_i}$, $\overline{P_i}$, and $\underline{P_i}$, $\overline{V_i}$, $\underline{V_i}$, $\overline{Q_i}$, $\underline{Q_i}$, are upper and lower boundaries of frequency, real power, voltage, reactive power of ith inverter.

The inverter output voltage v_{di}, v_{qi} are regulated to the reference voltage value v_{di}^*, v_{qi}^* determined by the droop controller. The voltage control loop is as follows:

$$\dot{\phi}_{di} = v_{di}^* - v_{di} \tag{7}$$

$$\dot{\phi}_{qi} = v_{qi}^* - v_{qi} \tag{8}$$

$$i_{ldi}^* = K_{PVi}(v_{di}^* - v_{di}) + K_{IVi}\phi_{di} \tag{9}$$

$$i_{lqi}^* = K_{PVi}(v_{qi}^* - v_{qi}) + K_{IVi}\phi_{qi} \tag{10}$$

The current control loop is as follows:

$$\dot{\gamma}_{di} = i_{ldi}^* - i_{ldi} \tag{11}$$

$$\dot{\gamma}_{qi} = i_{lqi}^* - i_{lqi} \tag{12}$$

$$v_{idi} = K_{PCi}(i_{ldi}^* - i_{ldi}) + K_{ICi}\gamma_{d,i} - \omega_i L_{fi} i_{lqi} \tag{13}$$

$$v_{iqi} = K_{PCi}(i_{lqi}^* - i_{lqi}) + K_{ICi}\gamma_{q,i} + \omega_i L_{fi} i_{ldi} \tag{14}$$

where ϕ_{di}, ϕ_{qi}, γ_{di}, and γ_{qi} are state variables of voltage and current control loops. K_{PVi}, K_{IVi}, K_{PCi}, and K_{ICi} are proportional and integral gains of voltage and current control loops. i_{ldi}^*, i_{lqi}^* are the reference generated by voltage control, which will be tracked by current control. i_{ldi}, v_{iqi} are the current and voltage measurement before the LC filter.

The differential equations for the output LC filter are as follows:

$$L_{fi}\dot{i}_{ldi} = -R_{fi} i_{ldi} + L_{fi}\omega_i i_{lqi} + v_{idi} - v_{di} \tag{15}$$

$$L_{fi}\dot{i}_{lqi} = -R_{fi} i_{lqi} - L_{fi}\omega_i i_{ldi} + v_{iqi} - v_{qi} \tag{16}$$

$$C_{fi}\dot{v}_{di} = C_{fi}\omega_i v_{qi} + i_{ldi} - i_{di} \tag{17}$$

$$C_{fi}\dot{v}_{qi} = -C_{fi}\omega_i v_{di} + i_{lqi} - i_{qi} \tag{18}$$

where R_{fi}, L_{fi} are the resistance and inductance of ith inverter.

2) *Network and Loads:* For a multi-inverter system, the interconnected variables of each inverter with the network and loads should be transferred between local *dq* frame and the common *DQ* frame. Specifically, the output voltage of the inverter is transferred to *DQ* frame by $V_{DQ} = \mathcal{T}(\delta) v_{dq}$. For the distribution line between bus i and bus k, the dynamic of the line current in *DQ* frame is represented as:

$$L_{ik}\dot{I}_{Dik} = -R_{ik} I_{Dik} + \omega_0 L_{ik} I_{Qik} + V_{Di} - V_{Dk} \tag{19}$$

$$L_{ik}\dot{I}_{Qik} = -R_{ik} I_{Qik} - \omega_0 L_{ik} I_{Dik} + V_{Qi} - V_{Qk} \tag{20}$$

where R_{ik}, L_{ik} are the resistance and inductance between bus i and k. ω_0 is a constant synchronous frequency.

An equivalent resistance-inductance (RL) load is considered at each bus in the systems. The dynamic of RL load connected at bus i in *DQ* frame can be expressed as:

$$L_{Li}\dot{I}_{LDi} = -R_{Li} I_{LDi} + \omega_0 L_{Li} I_{LQi} + V_{Di} \tag{21}$$

$$L_{Li}\dot{I}_{LQi} = -R_{Li} I_{LQi} - \omega_0 L_{Li} I_{LDi} + V_{Qi} \tag{22}$$

where R_{Li}, L_{Li} indicate equivalent resistance and inductance of RL load in bus i. As the current is balanced at each bus, thus the current injection by each inverter are $I_{Di} = I_{Dik} + I_{LDi}$, $I_{Qi} = I_{Qik} + I_{LQi}$. Then the current of the inverter can be transferred back to local *dq* frame by $i_{dq} = \mathcal{T}(\delta)^{-1} I_{DQ}$. Based on the power calculation in local *dq* frame, it can be obtained that:

$$P_i = 1.5(i_{di} v_{di} + i_{qi} v_{qi}) \tag{23}$$

$$Q_i = 1.5(i_{di} v_{qi} - i_{qi} v_{di}) \tag{24}$$

Then the calculated power P_i and Q_i are applied in the droop control in (3) and (4).

B. Small-Signal Model

The small-signal model is widely used to analyze the stability of inverter-based microgrids. The 5-order system by simplifying the inner control loops and LC filter dynamics in (7)-(18) is considered. The reduced order system still offers high accuracy for calculating the eigenvalues and evaluating the system stability [25]. By linearizing the above system around the operational or equilibrium point using Taylor expansion, the small-signal model can be obtained. The equilibrium point can be obtained by solving the differential equations of the full-order high-fidelity model. Therefore, by integrating the state equation of inverters, network, and loads, the small-signal model of the multi-inverter system can be obtained as:

$$\Delta \dot{x}_{sys} = A_{sys} \Delta x_{sys} \tag{25}$$

where $\Delta x_{sys} = [\Delta\delta, \Delta\omega, \Delta V, \Delta\widetilde{I}_D, \Delta\widetilde{I}_Q]^T$ is the state variable of the system. $\Delta\widetilde{I}_D = [\Delta I_{Line,D}, \Delta I_{Load,Q}]^T$, $\Delta\widetilde{I}_Q = [\Delta I_{Line,Q}, \Delta I_{Load,Q}]^T$. $I_{Line,D}, I_{Line,Q}, I_{Load,D}, I_{Load,Q}$ are vectors of line current and load current. An incidence matrix ∇^T of the power network is introduced, where $\nabla^T_{ij} = 1$ if current of jth line is injected to ith bus, $\nabla^T_{ij} = -1$ represents the current of jth line leaves ith bus. $x^s \in \{\delta^s, \omega^s, V^s, \widetilde{I}^s_D, \widetilde{I}^s_Q\}$ is the equilibrium points of the system, A_{sys} is detailed coefficient matrix of the system, which is given as follows

$$A_{sys} = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \Theta_{21} & -\omega_c\mathbf{1} & \Theta_{23} & \Theta_{24} & \Theta_{25} \\ \Theta_{31} & \mathbf{0} & \Theta_{33} & \Theta_{34} & \Theta_{35} \\ \Theta_{41} & \mathbf{0} & \Theta_{43} & -\omega_0 R X^{-1} & \omega_0\mathbf{1} \\ \Theta_{51} & \mathbf{0} & \Theta_{53} & -\omega_0\mathbf{1} & -\omega_0 R X^{-1} \end{bmatrix}$$

and

$$\begin{aligned}
\Theta_{21} &= -1.5\omega_c\mathcal{M}[-V^s_D \sin(\delta^s) I^s_D + V^s_D \cos(\delta^s) I^s_Q], \\
\Theta_{23} &= -1.5\omega_c\mathcal{M}[\cos(\delta^s) I^s_D + \sin(\delta^s) I^s_Q], \\
\Theta_{24} &= -1.5\omega_c\mathcal{M} V^s_D \nabla^T, \\
\Theta_{25} &= -1.5\omega_c\mathcal{M} V^s_Q \nabla^T, \\
\Theta_{31} &= -1.5\omega_c\mathcal{N}[V^s_D \cos(\delta^s) I^s_D + V^s_D \sin(\delta^s) I^s_D], \\
\Theta_{33} &= -\omega_c\mathbf{1} + 1.5\omega_c\mathcal{N}[\sin(\delta^s) I^s_D - \cos(\delta^s) I^s_Q], \\
\Theta_{34} &= -1.5\omega_c\mathcal{N} V^s_Q \nabla^T, \\
\Theta_{35} &= 1.5\omega_c\mathcal{N} V^s_D \nabla^T, \\
\Theta_{41} &= -\omega_0 X^{-1} \nabla V^s_D \sin(\delta^s), \\
\Theta_{43} &= \omega_0 X^{-1} \nabla \cos(\delta^s), \\
\Theta_{51} &= \omega_0 X^{-1} \nabla V^s_D \cos(\delta^s), \\
\Theta_{53} &= \omega_0 X^{-1} \nabla \sin(\delta^s).
\end{aligned}$$

where V^s_D, V^s_Q, I^s_D, I^s_Q, δ^s are equilibrium points in diagonal matrix form, which can be obtained from the time-domain simulation of the non-linear model presented in Section II.A. $\mathcal{M} = diag\{m_1, m_2, ..., m_N\}$, $\mathcal{N} = diag\{n_1, n_2, ..., n_N\}$ are a diagonal matrix of droop control gains. R and X are resistance and inductance matrices with network and loads. It is noted that this model is scalable according to the inverters, buses, and loads in the system. The derivation process of this model is omitted for brevity. The system contains $3N_{Inv} + 2N_{Load} + 2N_{Line}$ of states. N_{Inv}, N_{Load}, N_{Line} are the number of inverters, loads, and distribution lines.

III. Analysis of Destabilizing Attack and Defense on Inverter-based Microgrids

In the studied multi-inverter systems, the droop control gains of each IBR are adjustable. It can be changed to adapt to grid conditions, or dispatched by the system operator via communication systems [7]. In the meantime, the parameters of inner control loops are particularly designed for each inverter by the manufacturer, which is usually non-changeable. The *attack surface* of the multi-inverter systems are considered as these flexible parameters, such as droop control gains and their power set-points. The droop control gains will influence the stability of the system, while the power set points influence the equilibrium points. In this study, we mainly focus on destabilizing attacks by adjusting the droop control gains and their influence on the small-signal stability of the system.

A. Attack and Defense on Droop Gains

First, all the droop gains of IBRs in the inverter-based microgrids are separated into two sets. The attack set $\mathcal{V}_{att}$ contains droop gains which can be manipulated by attackers to destabilize the system. The defense set $\mathcal{V}_{def}$ contains droop gains which can be controlled by defenders to stabilize the system. Based on the attack and defense sets, the IBRs in the system can be separated into victim IBRs and defense IBRs. Therefore, the system under attack and defense can be formulated as

$$x(t+1) = f(x(t), u_{att}(t), u_{def}(t), t), \tag{26}$$

$$u_{att}(t) = h(x(t), t), \tag{27}$$

$$u_{def}(t) = g(x(t), t), \tag{28}$$

where u_{att} is the cyber-attack strategy target to the stability of the system. u_{def} is the defense strategy to maintain stability of the system. $u_{att} = 0$ or $u_{def} = 0$ means there is no cyber attack or defense control. The small-signal model of inverter-based microgrids with attack and defense on droop gains becomes

$$\Delta \dot{x}_{sys} = (A_{sys} + A_{att} + A_{def}) \Delta x_{sys} \tag{29}$$

where A_{att} is a matrix denoting the droop control gain change in victim IBRs. Specifically, the original droop gains m_i, n_i within the attack set $\mathcal{V}_{att}$ will be manipulated by $m_{att,i}$, $n_{att,i}$. That is to say all terms in $\mathcal{M}$ and $\mathcal{N}$ within the attack set will be changed as compared to the original system matrix A_{sys}. It will finally influence the small-signal stability of the multi-inverter system. To defend such attack, the system operators can design certain strategies to change the droop gains of defense IBRs. A_{def} is a matrix denoting the droop control gain change in defense IBRs. Specifically, the original droop gains m_i, n_i within the defense set $\mathcal{V}_{def}$ will be changed by $m_{def,i}$, $n_{def,i}$.

TABLE I
PARAMETERS OF A MICROGRID SYSTEM WITH 4 IBRS

Parameters	**Values**			
IBR (No.)	1	2	3	4
m_i (rad/W)	1×10^{-4}	1×10^{-4}	0.5×10^{-4}	0.5×10^{-4}
n_i (V/Var)	1×10^{-4}	1×10^{-4}	0.5×10^{-4}	0.5×10^{-4}
$[\underline{P_i}, \overline{P_i}]$ (kW)	[0, 20]	[0, 20]	[0, 20]	[0, 20]
$[\underline{Q_i}, \overline{Q_i}]$ (kVar)	[-20, 20]	[-20, 20]	[-20, 20]	[-20, 20]
Load (Bus No.)	1	2	3	4
R (Ω)	5.6	14.1	11.2	8.3
L (mH)	8.9	45	17.8	8.8
Line (No.)	1-2	2-3	3-4	
R (Ω)	0.16	0.32	0.24	
L (mH)	1.1	2.2	1.7	
LC filter (No. 1-4)	$C_f = 50\mu$F	$L_f = 5$mH	$R_f = 0.1\Omega$	
Inner control	K_{PV}=50	K_{IV}=250	K_{PC}=50	K_{IC}=500
Reference	ω^*=$100\pi rad$	V^*=$220\sqrt{2}V$	$\tau = 27$ ms	

Recall the definition of the eigenvalue and its eigenvectors:

$$A\phi_i = \lambda_i \phi_i, \tag{30}$$

$$\psi_i^T A = \psi_i^T \lambda_i. \tag{31}$$

where ϕ_i and ψ_i are the right and left eigenvectors of λ_i. The eigenvalue of A can be obtained by solving the determinant $det(A - \lambda_i I) = 0$. The negative value for the real part of λ_i indicates stable modes, while the zero value for marginally stable modes and the positive value for unstable modes.

The spectral abscissa of the system matrix A is the maximum real part of its eigenvalues, which can be presented as [26]:

$$\Lambda(A) = \max\{\mathrm{Re}\{\lambda_i\} : \det(A - \lambda_i I) = 0\}, \tag{32}$$

Here we further define $\widetilde{\Lambda}(A)$ as the spectral abscissa with non-zero imaginary part $\mathrm{Im}\{\lambda_i\} \neq 0$, considering the fact that the eigenvalues will not shift to the right plane when they are on the real axis.

The damping ratio ζ_i is defined as

$$\zeta_i = \frac{-\alpha_i}{\sqrt{\alpha_i^2 + \beta_i^2}} \tag{33}$$

where α_i and β_i are the real and imaginary parts of λ_i. It describes the attenuation of the system oscillations. In addition, the sensitivity of eigenvalue λ_i with respect to a parameter κ can be calculated by

$$\frac{\lambda_i}{\partial \kappa} = \psi_i^T \frac{\partial A}{\partial \kappa} \phi_i \tag{34}$$

It is noted that the calculation of eigenvalue involves non-convexity, as well as its associated factors including spectral abscissa, damping ratio, and sensitivity of eigenvalue regarding system parameters, which brings difficulty into related optimization problems [8], [27].

B. Analysis with a Microgrid Example

The destabilizing attack on droop gains aims to shift the eigenvalues of the original system A_{sys} into the unstable ones. A microgrid system with 4 IBRs is used as an example to show how droop gain change will influence the system's stability. The detailed parameters of the microgrid with 4 IBRs are shown in Table I. The system stability region of the small-signal model regarding frequency and voltage droop gains is shown in Fig. 2. The eigenloci under the change of m_i in the small-signal model is shown in Fig. 3. As shown in Fig. 2 (a), there are two dimensions in the stability region to be changed, i,e, frequency and voltage droop gains m_i, n_i. As shown in Fig. 2 (b), the changing of m_3 in IBR-3 will shift the stability region of IBR-4. It indicates the defender can change the droop gains of certain IBRs in order to stabilize the system under the attack of other IBRs. As shown in Fig. 3, by changing the m_i from 5×10^{-5} to 1×10^{-3} respectively, the mode of λ_{15} and λ_{16} will be moved towards to right plane. With a sufficient amount of manipulation of droop gains, the system will become unstable. The defender has the opposite goal, which aims to allocate all eigenvalue to the left plane. Besides, it can be found from Fig. 2 (a) that IBR-3 has the smallest stability region. It indicates the IBR-3 is the most vulnerable part under destabilizing attack, which should be well protected. In addition, in order to defend against an attack on certain droop gains, the defender should have more resources than the attacker, so that the eigenvalue can be shifted to desired regions.

Considering the attack and defense sets of IBRs, there are three general cases:
(a) The attack and defense sets have no intersection, i.e. $\mathcal{V}_{att} \cap \mathcal{V}_{def} = \varnothing$. There are two sub-conditions: i) The IBRs under attack and defense are different. In this condition, the defender can change m_i of other IBRs to make the destabilizing attack not successful, as shown in Fig. 2(b). ii) The attacker can only manipulate either frequency droop gain m_i or voltage droop gain n_i. This condition can be found when frequency/voltage droop gains have different communication channels and are dispatched separately. It can be illustrated by Fig. 2 (a), where the attacker can only change m_i or n_i. If the attacker manipulates m_i of certain IBR, the defender can adjust n_i to make the system operation point in the stability region. Similarly, if the attacker manipulates n_i of certain IBR, the defender can reduce m_i below a certain value.
(b) The attack and defense sets are equal, i.e. $\mathcal{V}_{att} = \mathcal{V}_{def}$. It happens when certain inverters are subject to attack, but the defender does not lose control ability of them. If the defender can adapt to the changes of the attackers, then the attack can be defended.
(c) The attack and defense sets have a partial intersection, i.e $\mathcal{V}_{att} \cap \mathcal{V}_{def} \neq \varnothing$. This is a more general condition as compared to cases (a) and (b), a mixed strategy can be taken by the defenders. Therefore, a proper method should be developed to find the combination of droop gains in the defense set.

C. Least Effort Attack

From victim IBRs, one can find the least effort attack with minimal changes of droop control gains. Thus, the least effort attack is defined as the attack which has minimal changes of $m_{att,i}$ and $n_{att,i}$ within attack set $\mathcal{V}_{att}$. The LEA can be represented as the problem below:

$$\min_{m_{att,i}, n_{att,i}} \sum_{i \in \mathcal{V}_{att}} (|m_{att,i}| + |n_{att,i}|) \tag{35}$$

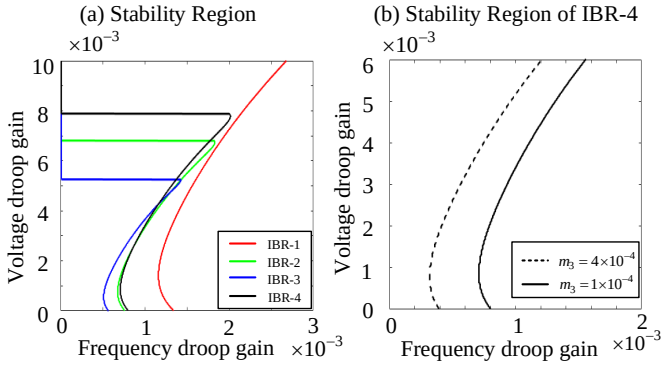

Fig. 2. (a) The stability region of the small-signal model regarding frequency and voltage droop gains of m_i and n_i. (b) The stability region of IBR-4 with respect to the change of frequency droop gain of IBR-3 m_3. The left region of the curve indicates the system is stable.

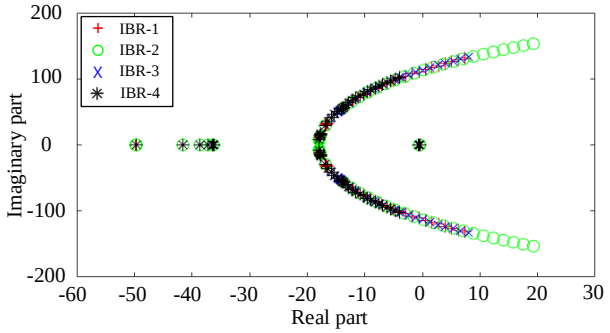

Fig. 3. The eigenloci of the small-signal model regarding the change of m_i. The range of m_i is from 5×10^{-5} to 1×10^{-3}. The other parameters are fixed with original values in Table I when m_i changes.

$$s.t. det(A_{sys} + A_{att} - \lambda_i I) = 0 \tag{36}$$
$$\Lambda(A_{sys} + A_{att}) > 0 \tag{37}$$
$$\underline{m_i} < m_i + m_{att,i} < \overline{m_i} \tag{38}$$
$$\underline{n_i} < n_i + n_{att,i} < \overline{n_i} \tag{39}$$

This problem is to find the argument of $m_{att,i}$ and $n_{att,i}$ that minimizes the above problem. (36) is the determinant for calculation of the system eigenvalues. (37) represents that the system spectral abscissa should be larger than zero, which leads to system unstable. (37) can be replaced by $\widetilde{\Lambda}(A_{sys} + A_{att}) = \Lambda^*_{att}$, if the attack aims to place the spectral abscissa into a specific value. Besides, λ_i or ζ_i can also be replaced into the constraints if the attack targets to specific eigenvalues and modes. Inequalities (38)–(39) impose upper and lower bounds on total droop gains of each IBR. They are the preset limits of the IBR which can not be violated.

The above formulation is for the static LEA, which does not consider the system change. Considering this problem in a dynamic environment, it is equivalent to find the optimal attack strategy $u^*_{att,t}$ considering dynamic system in (26). As the small-signal model can describe the system stability at each time interval. Therefore, by considering time interval t into the above LEA problem, the sequential or dynamic LEA can be formulated. Both LEA problems contain the calculation of the eigenvalues and spectral abscissa of the system under attack. As the eigenvalue sensitivity in (34) of the studied system is highly non-linear, it is hard to estimate the final value based on the original condition. Besides, in real operation conditions, there will be parameter variations in the inverter-based microgrids.

D. Defense Strategy

To defend dynamic LEA on droop gains in inverter-based microgrids, a defense strategy can be designed to change the droop gain by $m_{def,i}$ and $n_{def,i}$ within defense set $\mathcal{V}_{def}$. Therefore, the defense problem can be represented as

$$\min_{m_{def,i}, n_{def,i}} \sum_{t\in T} \sum_{i\in\mathcal{V}_{def}} (|m_{def,i,t}| + |n_{def,i,t}|) \tag{40}$$

$$s.t. det(A_{sys,t} + A_{att,t} + A_{def,t} - \lambda_i I) = 0 \tag{41}$$
$$\Lambda(A_{sys,t} + A_{att,t} + A_{def,t}) < 0 \tag{42}$$
$$\underline{m_i} < m_{i,t} + m_{att,i,t} + m_{def,i,t} < \overline{m_i} \tag{43}$$
$$\underline{n_i} < n_{i,t} + m_{att,i,t} + m_{def,i,t} < \overline{n_i} \tag{44}$$

This defense problem is similar to the above dynamic LEA problem. It aims to stabilize the system by adjusting the droop gains. (42) can be replaced by $\widetilde{\Lambda}(A_{sys,t} + A_{att,t} + A_{def,t}) = \Lambda^*_{def}$, if the defender aims to place the spectral abscissa into a specific value.

Again, it involves the calculation of eigenvalues and system non-linear dynamics, which makes the problem hard to be dealt with. In the next section, we propose to use deep reinforcement learning to obtain the online optimal solution for these attack and defense problems.

IV. Attack and Defense by Adversarial Deep Reinforcement Learning

In this section, the dynamic LEA and its defense are presented in detail. Firstly, the dynamic LEA is formulated into a MDP. Then robust defense problem under such an attack is formulated as an AMDP. The training and implementation framework is demonstrated in Fig. 4. As shown in Fig. 4, the reduced order small-signal model will calculate the eigenvalue of the system at each time interval during the offline training stage. The system equilibrium points are obtained from the high-fidelity model. In the online implementation stage, both the small-signal model and high-fidelity model can be applied, which function to simulate the system trajectory and calculate the system eigenvalue. The system eigenvalue is obtained by the attack agent and defense agent to train the optimal policy. The attack agent and defense agent have opposite objectives, i.e. destabilize and stabilize the system. Then the droop gains in the attack and defense sets will be updated in the next time interval. Next, we need to transfer the attack and defense problems into MDP and AMDP forms, so that they can be handled by DRL methods. In this paper, TD3 is adopted to train DNN to learn the optimal attack and defense policy, which aims to find the optimal combination of droop gain change within the attack set and defense set. The deterministic policy can give a smoother output than stochastic policy in the studied problem. The TD3 as an extension method of DDPG, addresses the sub-optimal policies generated by the value function overestimation of DDPG. More details of TD3 can be found in its fundamental work [28].

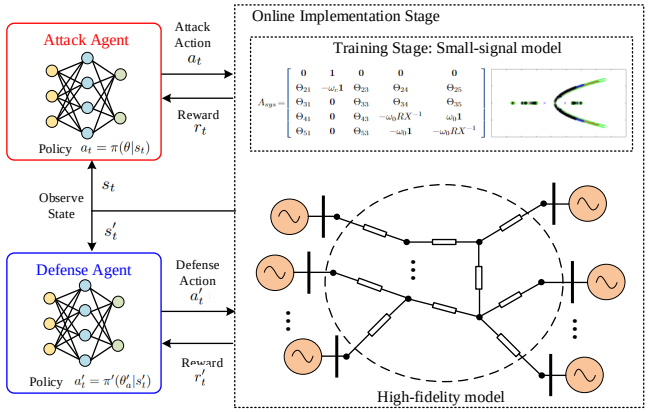

Fig. 4. The training and implementation framework of the proposed framework. The TD3 is used to accomplish the attack and defense task in inverter-based systems.

A. Dynamic LEA by Deep Reinforcement Learning

In this paper, the attack problem is formulated as the MDP defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$, where $\mathcal{S}$ presents a set of states from the environment, $\mathcal{A}$ is a set of actions, $\mathcal{P}$ is a set of transition probability function, $\mathcal{R}$ is a set of immediate rewards. The attack agent will learn an optimal control policy through interaction with the environment. It is noted that only the concerning states in the environment are received by the attack agent.

At each time step t, the attack agent will receive a state $s_t \in \mathcal{S}$ from the current state of the environment. Then the agent will generate an action $a_t \in \mathcal{A}$, which controls the environment into a new state s_{t+1}. The action is generated by a policy $\pi : \mathcal{S} \mapsto \mathcal{A}$ such that $a_t = \pi(s_t)$. In each time step, the agent will also receive a reward $r_t \in \mathcal{R}$, which is a function of the state and the action, i.e., $r : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{R}$. The transition between the environment states can be modelled by the transition probability function $\mathcal{P}(s_t, a_t, \sigma)$, where σ represents the uncertainty in the environment.

The goal of the attack agent is to learn an optimal policy π^* that maximizes the accumulated expected discounted reward $J(\pi) = \mathbb{E}(\sum_{t=0}^{T} \gamma^t r_t)$. Here T is the episode length and $\gamma \in [0, 1]$ is a discount factor. To estimate the expected discounted reward by taking action a_t following policy π in state s_t, the Q value function is defined i.e. $Q^\pi(s_t, a_t) = \mathbb{E}_\pi[J(\pi) \,|\, s_0 = s_t, a_0 = a_t.]$. In the actor and critic structure based DRL methods such as TD3, the Q value function is estimated by one or two DNNs as $Q_k(\theta_{c,k}|s_t, a_t)$. Besides, the actor is also based on DNN to generate the deterministic policy $a_t = \pi(\theta_a|s_t)$. $\theta_{c,k}$ and θ_a are the parameters of DNNs for critics and actor. In this paper, TD3 is used for training the DNN to solve the formulated MDP.

1) *State*: In the studied dynamic LEA problem, the agent will receive certain states from the environment. At time step t, the measured state is represented by:

$$s_t = \{\widetilde{\Lambda}(A_{sys,t} + A_{att,t}), \epsilon_{att}\} \in \mathcal{S}, \tag{45}$$

where $\epsilon_{att} = |\widetilde{\Lambda}(A_{sys} + A_{att}) - \Lambda^*_{att}|$ is the error to attacker targeted spectral abscissa, which can be calculated from the small-signal model.

2) *Action*: In this paper, the actions generated by the agent is defined as droop gains in the attack set. It aims to find the least effort attack path for system instability. Therefore, the action set of attack agent is defined as:

$$a_t = \{m_{att,i,t}, n_{att,i,t}\} \in \mathcal{A} \tag{46}$$

where $m_{att,i,t}$, $n_{att,i,t}$ refer to droop gains to be changed in the attack set. The total droop gains should be within the limits $[\underline{m_i}, \overline{m_i}]$, and $[\underline{n_i}, \overline{n_i}]$, as defined in (38), (39).

3) *State Transition*: The system state transition is governed by $s_{t+1} = \mathcal{P}(s_t, a_t, \sigma_t)$, which is determined jointly by current state s_t, agent action a_t and environment uncertainty σ_t. σ_t refers to the system uncertainty e.g. parameters of droop gains and RL value of loads. The agent will gradually learn the characteristics from the data sources of the environment.

4) *Reward*: The reward function r_t is used to evaluate the performance of action a_t at state s_t. The reward function is defined that the attack problem in (35)-(39) can be solved considering the stochastic environment. Thus the reward can be defined as:

$$r_t = r_{1,t} + r_{2,t}, \tag{47}$$

$$r_{2,t} = -\sum_{i \in \mathcal{V}_{att}} (|m_{att,i,t}| + |n_{att,i,t}|). \tag{48}$$

$$r_{1,t} = -|\widetilde{\Lambda}(A_{sys,t} + A_{att,t}) - \Lambda^*_{att}|, \tag{49}$$

The reward function contains two parts $r_{1,t}$ and $r_{2,t}$. The first part aims to find the minimal sum of $m_{att,i}$ and $n_{att,i}$ in the attacks set, as defined in (35) The second part is to shift the spectral abscissa with non-zero imaginary part $\widetilde{\Lambda}$ to a desired positive value Λ^*_{att}, as defined in (37).

B. Robust Defense by Adversarial Reinforcement Learning

Given an agent with attack policy π, we wish to learn a policy π' to defend such an attack. This can be achieved by solving the AMDP problem $(\mathcal{S}', \mathcal{A}', \mathcal{P}', \mathcal{R}')$. Similar to the attack problem as MDP, $\mathcal{S}'$ is the set of system states including the attack agent. $\mathcal{A}'$ is the set of all available defense actions. $\mathcal{P}' : \mathcal{S} \times \mathcal{A} \times \mathcal{A}'$ is the transition probability under attack policy π and defense policy π'. $\mathcal{R}'$ is the reward function of defense agent, which can be chosen as the opposite reward of the attack agent. The defense agent seeks to minimize the expected reward with the attack agent as $\min_{\pi'} \max_{\pi} J'(\pi', \pi)$. As the iterative training of attack policy π and defend policy π' converges slowly and does not provide greater robustness [29]. Here we adopt the alternative way which is to fix the attack policy π when training the defense policy $a'_t = \pi'(\theta'_a|s'_t)$.

1) *State*: The defense agent receives similar states from the environment. At time step t, the measured state is represented by:

$$s'_t = \{\widetilde{\Lambda}(A_{sys,t} + A_{att,t} + A_{def,t}), \epsilon_{def}\} \in \mathcal{S}', \tag{50}$$

where $\epsilon_{def} = |\widetilde{\Lambda}(A_{sys,t} + A_{att,t} + A_{def,t}) - \Lambda^*_{def}|$ is the error to defender targeted spectral abscissa.

2) *Action*: In this paper, the actions generated by the defense agent are defined as droop gains in the defense set. It aims

to find a policy that stabilizes the system. The action set of defense agent is defined as:

$$a'_t = \{m_{def,i,t}, n_{def,i,t}\} \in \mathcal{A}' \tag{51}$$

where $m_{def,i,t}$, $n_{def,i,t}$ refer to droop gains to be changed in the defense set. The total droop gain should be within the limits $[\underline{m_i}, \overline{m_i}]$, and $[\underline{n_i}, \overline{n_i}]$, as given in (43), (44).

3) *State Transition*: The system state transition is governed $s'_{t+1} = \mathcal{P}'(s'_t, a_t, a'_t, \sigma_t)$, which is determined jointly by current state s'_t, attacker agent action a_t, defender agent action a'_t and environment uncertainty σ_t.

4) *Reward*: The reward function r'_t is used to evaluate the performance of action a'_t at state s'_t. The reward function is defined that the defense problem in (40)-(44) can be solved. Thus the reward of defense agent can be defined as:

$$r'_t = r'_{1,t} + r'_{2,t}, \tag{52}$$

$$r'_{2,t} = -\sum_{i \in \mathcal{V}_{def}} (|m_{def,i,t}| + |n_{def,i,t}|). \tag{53}$$

$$r'_{1,t} = -|\widetilde{\Lambda}_(A_{sys,t} + A_{att,t} + A_{def,t}) - \Lambda^*_{def}|, \tag{54}$$

The reward function contains two parts $r'_{1,t}$ and $r'_{2,t}$. It is to find the minimal change of $m_{def,i}$ and $n_{def,i}$ in the defense set to shift the $\widetilde{\Lambda}$ to a desired negative value Λ^*_{def}, as given in (40) and (42). Based on the TD3 algorithm, the above AMDP can be solved. It is achieved by adversarial training of the attack agent with the defense agent. The process of adversarial training is summarized in Algorithm 1.

Algorithm 1 Adversarial Training

1: Import trained attack agent with policy $\pi \sim \theta_a$.
2: Initialize defense agent with randomized actor network $\pi' \sim \theta'_a$ and critic networks $Q'_1 \sim \theta'_{c,1}$,$Q'_2 \sim \theta'_{c,2}$. The target networks are of the same size.
3: Set training hyperparameters of TD3 as in Table II
4: **for** $episode$ = 1 to M **do**
5: Initialize state s'_1 and droop gains m_i, n_i within a range.
6: **for** t = 1 to T **do**
7: Determine action a'_t by policy $\pi'(\theta'_a|s'_t)$
8; Take action a'_t, get reward r'_t and observe the next state s'_{t+1}
9: Store the transition s'_t, a'_t, r'_t into the replay buffer R.
10: **end for**
11: A mini-batch of m instances is randomly sampled from R.
12: Update the actor and critic networks parameters with policy gradient by TD3.
13: **end for**

V. Results

A. Test Setup

1) *Test systems*: To evaluate the performance of the proposed method, two test systems, including 4-IBR microgrids and IEEE 123 bus system with 10 IBRs, are considered. The structure of the 4-IBR microgrid is shown in Fig. 1. In this system, 4 IBRs are connected to the microgrid via inverters, and 4 RL loads are connected at each bus. The parameters of the system are given in Table I. Besides, the performance of the proposed method is also evaluated under a large-scale system by using IEEE 123-bus systems.

TABLE II
Hyperparameters of TD3 Algorithm

Hyperparameters	**Values**
Experience buffer length	1×10^6
Minibatch size	256
Discount factor	0.99
Actor learning rate	1×10^{-4}
Critics learning rate	1×10^{-3}
Optimizer	Adam
Policy updating frequency	2
Target smooth factor	1×10^{-3}
Noise standard deviation	0.1
Noise standard deviation decay rate	1×10^{-4}

2) *Training Setup*: The TD3 algorithm is used to solve the MDP and AMDP as well as train the attack and defense agents. The hyperparameters of TD3 used for attack and defense agents training are presented in Table II. The time interval between two consecutive steps $\Delta t = 0.1$ s. The training is performed on a laptop with 3.00GHz Intel i7-1185g7 CPU and 16GB RAM. DNNs are initialized with random weights and biases, which include an actor network and double critic networks. All actor and critic networks have two hidden layers with 100 and 50 units. The number of neurons at input and output layers vary according to the specific problems. The *ReLU* activation function is used for all hidden layers of actor and critic networks. The *Tanh* activation function is subsequently applied to the output of the actor network.

3) *Hyperparameter Selection*: In this paper, typical values are selected as in Table II to ensure modest training performance. Some suggestions for hyperparameter selection are as follows. The experience buffer stores past transitions for training. A large value can store more diverse transitions, but also requires more memory. The batch size determines the number of transitions to be used in each iteration of the training process. A large batch size means that more transitions are used in each iteration. It can increase the accuracy of the gradients computed during training, but also requires more computational resources. The discount factor determines the importance of future rewards in the calculation of the expected return. A large discount factor can lead to a focus on long-term rewards, while a small discount factor can lead to a focus on short-term rewards. The learning rate determines how quickly the model updates its weights. A large learning rate may lead to instability, while a small learning rate may result in slow convergence. The exploration noise is added to the actions produced by the actor network to encourage the agent to explore new actions and states. The scale of the exploration noise can affect the balance between exploration and exploitation.

B. Case 1: Least Effort Attack

First, the basic case that 4-IBR microgrids subjected to dynamic LEA is studied. It is considered that IBR-2 and IBR-4 is under attack, and the attack set contains $m_{att,2}$, $n_{att,2}$, $m_{att,4}$ and $n_{att,4}$. In the training stage, the droop gains of each inverter are randomly initialized among ranges

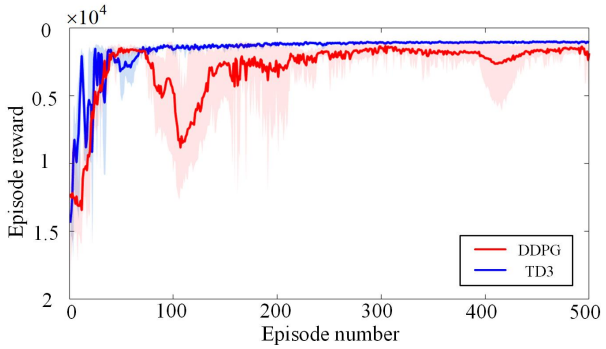

Fig. 5. The average reward and episode reward during the training stage of the attack agent.

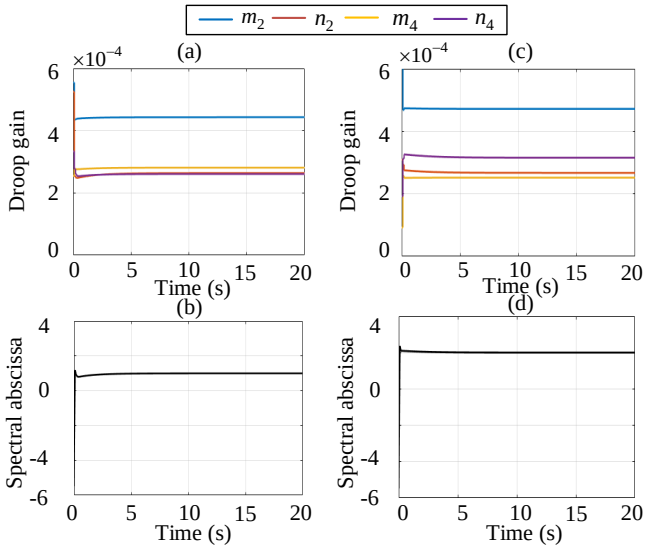

Fig. 6. The droop gains of IBR-2 and IBR-4 and spectral abscissa of the microgrids.

$(m_i, n_i \in [1.5\times10^{-4}, 3\times10^{-4}])$ in each episode, which makes the agent robust to parameter uncertainty. In Fig.5, the episode reward and average reward of 20 steps during the training are presented, and DDPG and TD3 methods are compared. The TD3 has better convergence and accumulated reward than DDPG for the studied problem as shown in Fig.5. Therefore, in the rest of this paper, TD3 is adopted by both attack and defense agents in the training. As the training progresses, the average and episode rewards increase and gradually become converged. The training shows asymptotic convergence within 500 episodes.

After the training is completed, the actor network can be extracted and used to find the dynamic LEA in a real-time environment. It is considered that the system operated at $m_i = 2.5\times10^{-4}$ and $n_i = 2.5\times10^{-4}$ when the simulation start. The droop gains of IBR-2 and IBR-4 and spectral abscissa of the system are shown in Fig. 6 (a)-(d). Fig. 6 (a) shows that the attack agent adjusts the droop gains into $m_2 = 4.43\times10^{-4}$, $n_2 = 2.64\times10^{-4}$, $m_4 = 2.81\times10^{-4}$ and $n_4 = 2.61\times10^{-4}$. As result, the spectral abscissa of the system is changed from -5.5 to 1 in Fig. 6 (b). Besides, we also set $\Lambda^*_{att} = 2$, where another group of droop gains can be found, as shown in Fig.6 (c) and (d). Fig. 6 (c) shows that the attack agent adjusts the

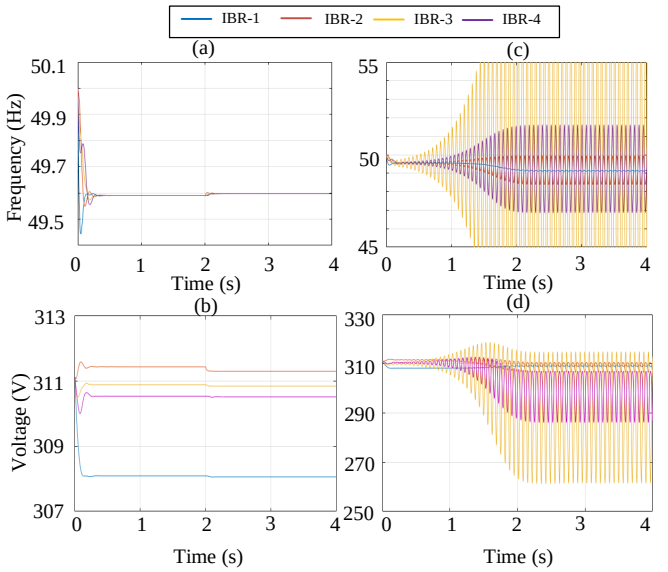

Fig. 7. The system frequency and voltage trajectories with and without attack agent and the attacker targeted spectral abscissa is 1.

droop gains into $m_2 = 4.72\times10^{-4}$, $n_2 = 2.67\times10^{-4}$, $m_4 = 2.52\times10^{-4}$ and $n_4 = 3.16\times10^{-4}$. The system frequency and voltage trajectories with and without attack agents are shown in Fig. 7. As shown in Fig.7 (a) and (b), the system under initial droop gains will operate stably. However, the inclusion of attack agent on the system will lead to the system instability in Fig.7 (c) and (d). The time-domain simulation with high-fidelity system validates that the attack targeted to maximum eigenvalue eventually destabilizes the system.

C. Case 2: Robust Defense Strategy

In the second case, the defense agent is added into the system of the basic case. The defense agent can change droop gains of IBR-2 and IBR-3, where the defense set contains $m_{def,2}$, $n_{def,2}$, $m_{def,3}$ and $n_{def,3}$ and has intersection with attack set. After adversarial training with attack agent, the actor network of defense agent can be deployed to defend such attack. The actions of droop gain changes from attack and defense agents, as well as spectral abscissa of the system, are shown in Fig. 8. In Fig. 8 (a) and (b), it can be found that certain droop gains change (e.g. droop gains of IBR-4) faster than others, and the defense droop changes slower than the attack droop. The underlying reason should be the defense droop are more sensitive to system eigenvalue or spectral abscissa than attack droop. Therefore, the defense agent can change them slowly to deal with the rapid change of droop by the attack agent. Besides, the attack agent and defense agent will gradually reach an equilibrium. The attack agent will also change its output as compared to Case-1. After the adversarial training against the attack agent, the defense agent is capable to bring the system spectral abscissa to -2 to make the system stable.

The system eigenvalue before attack, after attacks, and with defense are demonstrated in Fig. 9. It can be found that the system critical eigenvalue shifts to the right plane after attack. The imaginary part changed from -5.4 to 2. After the defense

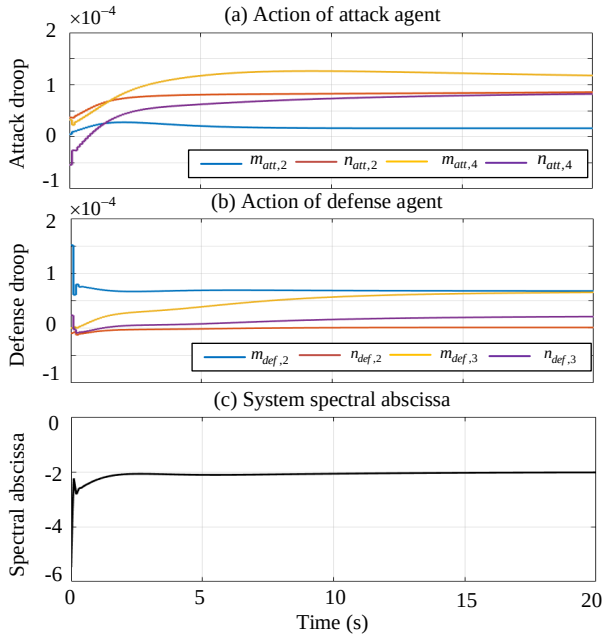

Fig. 8. The actions of droop gain changes from attack and defense agents as well as spectral abscissa of the microgrid.

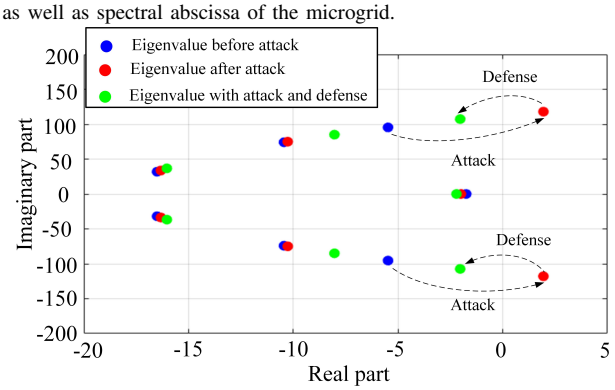

Fig. 9. The system eigenvalue of before attack, after attacks, and with both attack and defense.

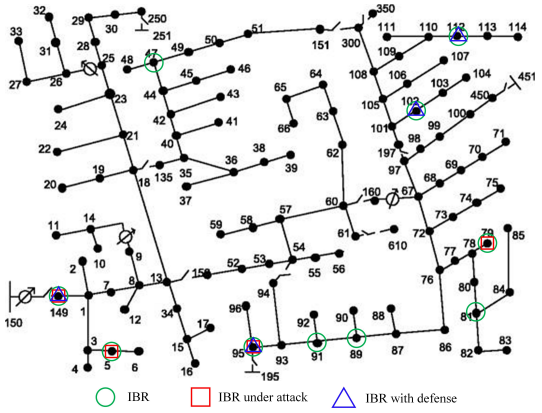

Fig. 10. The IEEE 123-bus system used for scalability test.

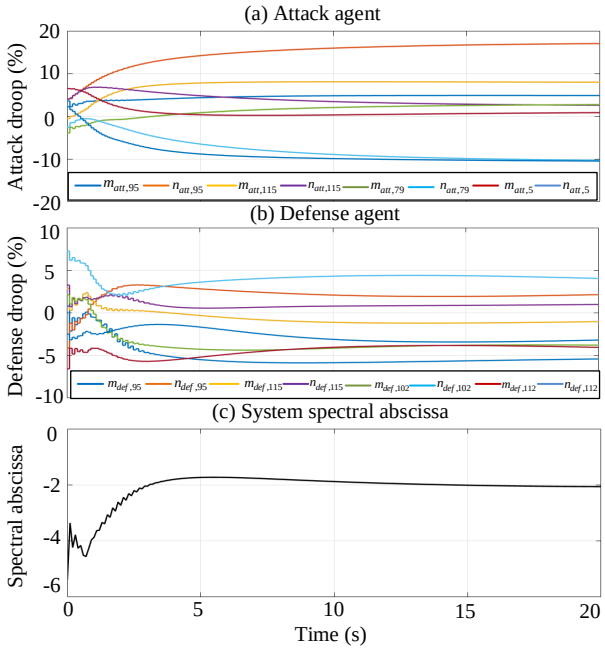

Fig. 11. The droop gain changes by attack and defense agents as well as spectral abscissa of the microgrid.

agent involves, the imaginary part can shift back to -2. In the shifting process, the proposed method will keep minimal change of droop gains.

D. Case 3: Scalability Test in IEEE 123-bus System

In the third case, the scalability of the proposed attack and defense framework is tested in a modified IEEE 123-bus system, where details can be found in [25]. The system topology is shown in Fig. 10. The IBRs are located at bus {95, 149, 79, 5, 102, 112, 81, 91, 89, 47}. The IBR at bus {95, 149, 79, 5} is under attack, while the IBRs at bus {95, 149, 102, 112} is under defense. The droop gain changes by attack and defense agents in p.u are shown in Fig. 11 (a) and (b). As droop gains in defense set are more sensitive to system stability than in attack set. Therefore, the defense agent can find a dynamic combination of droop gains to ensure the system's small-signal stability. Besides, as shown in Fig. 11 (c), the defense agent is capable to maintain the system spectral abscissa to -2.

The trace of eigenvalues during the simulation is shown in the 2D-plot of Fig. 12, while the trace of critical eigenvalues is shown in the 3D-plot of Fig. 13. As shown in Fig. 12, it can be found that the eigenvalues with the largest real part are shifted from the red star point to the green star point. The system special abscissa is changed from -5.38274 to -2.06296. It aligns with the time domain results shown in Fig. 11 (c). The system critical eigenvalues are considered as eigenvalues with the largest real part. As shown in Fig. 13, λ_{19}, λ_{20} are the system critical eigenvalues (blue circle) during 0s-1.9s. λ_{33}, λ_{34} are the system critical eigenvalues (red asterisk) during 2s-20s. This result also aligns with previous findings.

VI. Conclusions

In this paper, the problem of destabilizing attacks on droop gains in inverter-based microgrids is studied, and the data-driven destabilizing attack and robust defense strategy are proposed. Firstly, the full-order model and linearized reduced-order small-signal model of typical multi-inverter systems are derived. Then the destabilizing attack on the droop gains and its defense strategy is analyzed. Finally, a deep reinforcement

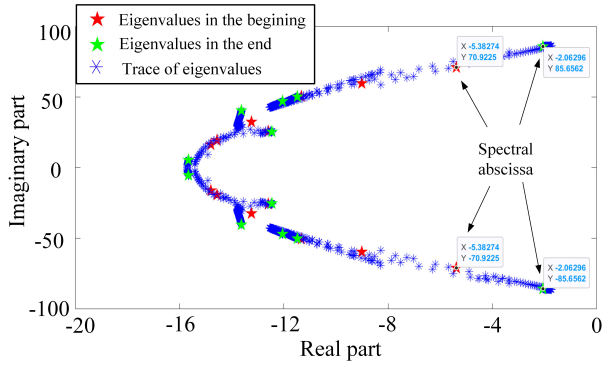

Fig. 12. The trace of system eigenvalues during the simulation in Case 3. The system special abscissa is changed from -5.38274 to -2.06296.

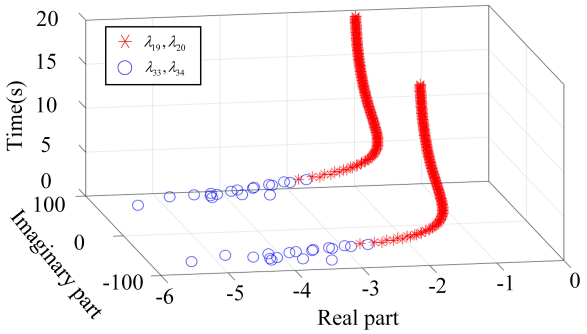

Fig. 13. The trace of system critical eigenvalues during the simulation in Case 3. The blue circle is for λ_{33}, λ_{34}. The red asterisk is for λ_{19}, λ_{20}.

learning approach TD3 is proposed to find the least effort attack path of this system and obtain the robust defense strategy. The simulation test results validate the effectiveness of the proposed method. It is found that the proposed method can determine the optimal combination of droop gains with attack and defense sets. The system spectral abscissa will be shifted to the targeted position by using the proposed method. The defense strategy obtained by adversarial training has robustness against the destabilizing attack. The test on IEEE 123 bus system validates the scalability of the proposed approach.

REFERENCES

[1] Achieving net zero: mapping the growth of the UK's energy transition , REA, UK. [Online Available]: https://www.r-e-a.net/wp-content/uploads/2020/03/FINAL-REview-2020.pdf.

[2] S. Mallapaty, "How China could be carbon neutral by mid-century." *Nature*, vol. 586, no. 7830, 22 Oct. 2020, pp. 482+.

[3] G. Hug and J. A. Giampapa, "Vulnerability Assessment of AC State Estimation With Respect to False Data Injection Cyber-Attacks," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1362-1370, Sept. 2012.

[4] P. Li, Y. Liu, H. Xin and X. Jiang, "A Robust Distributed Economic Dispatch Strategy of Virtual Power Plant Under Cyber-Attacks," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4343-4352, Oct. 2018.

[5] R. Tan, et al., "Modeling and mitigating impact of false data injection attacks on automatic generation control," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 7, pp. 1609-1624, July 2017.

[6] C. L. DeMarco, "Design of predatory generation control in electric power systems," *Proceedings of the Thirty-First Hawaii International Conference on System Sciences*, vol. 3. pp. 32–38. 1998.

[7] C. Roberts, U. Markovic, D. Arnold and D. S. Callaway, "Malicious Control of an Active Load in an Islanded Mixed-Source Microgrid," *IEEE Madrid PowerTech*, 2021.

[8] S. Amini, F. Pasqualetti and H. Mohsenian-Rad, "Dynamic Load Altering Attacks Against Power System Stability: Attack Models and Protection Schemes," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2862-2872, July 2018.

[9] C. Chen, Y. Chen, K. Zhang, M. Ni, S. Wang and R. Liang, "System Redundancy Enhancement of Secondary Frequency Control Under Latency Attacks," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 647-658, Jan. 2021.

[10] O. A. Beg, T. T. Johnson and A. Davoudi, "Detection of False-Data Injection Attacks in Cyber-Physical DC Microgrids," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2693-2703, Oct. 2017.

[11] Y. Wang, S. Mondal, C. Deng, K. Satpathi, Y. Xu and S. Dasgupta, "Cyber-Resilient Cooperative Control of Bidirectional Interlinking Converters in Networked AC/DC Microgrids," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 10, pp. 9707-9718, Oct. 2021.

[12] C. Deng, F. Guo, C. Wen, D. Yue and Y. Wang, "Distributed Resilient Secondary Control for DC Microgrids Against Heterogeneous Communication Delays and DoS Attacks," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 11, pp. 11560-11568, Nov. 2022.

[13] J. Liu, Y. Du, S. Yim, X. Lu, B. Chen and F. Qiu, "Steady-State Analysis of Microgrid Distributed Control Under Denial of Service Attacks," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 9, no. 5, pp. 5311-5325, Oct. 2021.

[14] J. Duan et al., "Deep-Reinforcement-Learning-Based Autonomous Voltage Control for Power Grid Operations," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 814-817, Jan. 2020.

[15] Z. Yan and Y. Xu, "Data-Driven Load Frequency Control for Stochastic Power Systems: A Deep Reinforcement Learning Method With Continuous Action Search," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1653-1656, March 2019.

[16] D. Qiu, Y. Wang, T. Zhang, M. Sun and G. Strbac, "Hybrid Multi-Agent Reinforcement Learning for Electric Vehicle Resilience Control Towards a Low-Carbon Transition," *IEEE Transactions on Industrial Informatics*, doi: 10.1109/TII.2022.3166215.

[17] X. Chen, G. Qu, Y. Tang, S. Low and N. Li, "Reinforcement Learning for Selective Key Applications in Power Systems: Recent Advances and Future Challenges," *IEEE Transactions on Smart Grid*, doi: 10.1109/TSG.2022.3154718.

[18] A. J. Abianeh, Y. Wan, F. Ferdowsi, N. Mijatovic and T. Dragičević, "Vulnerability Identification and Remediation of FDI Attacks in Islanded DC Microgrids Using Multiagent Reinforcement Learning," *IEEE Transactions on Power Electronics*, vol. 37, no. 6, pp. 6359-6370, June 2022.

[19] H. Zhang, D. Yue, C. Dou and G. P. Hancke, "Resilient Optimal Defensive Strategy of Micro-Grids System via Distributed Deep Reinforcement Learning Approach Against FDI Attack," *IEEE Transactions on Neural Networks and Learning Systems*, doi: 10.1109/TNNLS.2022.3175917.

[20] P. Chen, S. Liu, B. Chen and L. Yu, "Multi-Agent Reinforcement Learning for Decentralized Resilient Secondary Control of Energy Storage Systems Against DoS Attacks," *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 1739-1750, May 2022.

[21] H. Liu and W. Wu, "Two-Stage Deep Reinforcement Learning for Inverter-Based Volt-VAR Control in Active Distribution Networks," *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2037-2047, May 2021.

[22] L. Omnes, A. Marot, and B. Donnot, "Adversarial training for a continuous robustness control problem in power systems," in 2021 IEEE Madrid PowerTech, 2021, pp. 1–6.

[23] N. Pogaku, M. Prodanovic and T. C. Green, "Modeling, Analysis and Testing of Autonomous Operation of an Inverter-Based Microgrid," *IEEE Transactions on Power Electronics*, vol. 22, no. 2, pp. 613-625, March 2007.

[24] P. Vorobev, P.-H. Huang, M. Al Hosani, J. L. Kirtley, and K. Turitsyn, "High-fidelity model order reduction for microgrids stability assessment," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 874–887, 2017.

[25] A. Gorbunov, J. C. H. Peng, J. W. Bialek and P. Vorobev, "Identification of Stability Regions in Inverter-Based Microgrids," *IEEE Transactions on Power Systems*, vol. 37, no. 4, July 2022.

[26] G. K. Dill and A. S. e Silva, "Robust Design of Power System Controllers Based on Optimization of Pseudospectral Functions," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1756-1765, May 2013.

[27] A. Venkatraman, U. Markovic, D. Shchetinin, E. Vrettos, P. Aristidou and G. Hug, "Improving Dynamic Performance of Low-Inertia Systems Through Eigensensitivity Optimization," *IEEE Transactions on Power Systems*, vol. 36, no. 5, pp. 4075-4088, Sept. 2021,

[28] S. Fujimoto, H. Hoof, and D. Meger. "Addressing function approximation error in actor-critic methods." *International Conference on Machine Learning*, 2018.

[29] A. Gleave, M. Dennis, C. Wild, N. Kant, S. Levine, and S. Russell. "Adversarial policies: Attacking deep reinforcement learning." *International Conference on Learning Representations*, 2020.

Yu Wang (S'12-M'17) received the B.Eng. degree in Electrical Engineering and Automation from Wuhan University, China in 2011, and the M.Sc. and Ph.D. degree in Power Engineering from Nanyang Technological University, Singapore in 2012 and 2017, respectively. Currently, he is a Professor at the School of Electrical Engineering, Chongqing University. He was a Marie Skłodowska-Curie Individual Fellow at Control & Power Group, Imperial College London. His research interests include microgrid control and stability, power system operation and control, and cyber-physical systems.

Bikash C. Pal (M'00-SM'02-F'13) received B.E.E. (with honors) degree from Jadavpur University, Calcutta, India, M.E. degree from the Indian Institute of Science, Bangalore, India, and Ph.D. degree from Imperial College London, London, U.K., in 1990, 1992, and 1999, respectively, all in electrical engineering. Currently, he is a Professor in the Department of Electrical and Electronic Engineering, Imperial College London. His current research interests include renewable energy modelling and control, state estimation, and power system dynamics. He is Vice President Publications, IEEE Power & Energy Society. He was Editor-in-Chief of IEEE Transactions on Sustainable Energy (2012-2017) and Editor-in-Chief of IET Generation, Transmission and Distribution (2005-2012) and is a Fellow of IEEE for his contribution to power system stability and control.