Secure Distributed Control for Demand Response in Power Systems Against Deception Cyber-Attacks With Arbitrary Patterns

Shaohua Yang, *Student Member, IEEE*, Keng-Weng Lao, *Senior Member, IEEE*, Hongxun Hui, *Member, IEEE*, and Yulin Chen, *Member, IEEE*

***Abstract*—Demand response (DR) is a crucial component of power systems that can offer operating reserves by utilizing flexible loads on the demand side. With the development of communication, information, and control technologies, DR has formed a cyber-physical system. While deep cyber-physical coupling improves the performance of DR, it also introduces cyber-security threats, i.e., deception cyber-attacks (DCAs), which can lead to DR out-of-control and thus threaten the power system's safe operation. To this end, this paper proposes a secure distributed control to safeguard DR against DCAs. First, a cyber-physical DR community is developed based on a distributed control framework for offering operating reserve to power systems. In addition, considering different patterns, the impacts of DCAs on DR are quantified, revealing that different attack patterns can lead to various adverse consequences on DR, such as power deviation, delayed response, power fluctuation, etc. Furthermore, an anti-attack secure distributed control is developed for DR to counteract against arbitrary DCAs. In addition, rigorous proof based on Lyapunov theorem demonstrates that the proposed control can ensure the stability and convergence for the DR power regulation required by power systems, despite arbitrary DCAs. Finally, case studies validate the efficacy of the proposed control method.**

***Index Terms*—Demand response, deception cyber-attack, arbitrary attack pattern, secure distributed control, Lyapunov theorem.**

I. Introduction

With growing concerns about climate change caused by greenhouse gas emissions, the supply side of power systems is undergoing a shift towards renewable energies (RENs), e.g., photovoltaics and wind turbines [1], [2]. As per the 2023 statistics, the world added nearly 295 GW of RENs last year, accounting for an unprecedented 83% of total power additions globally [3]. However, RENs are intermittent and stochastic in nature; hence, grid-connected RENs can bring power fluctuations to power systems [4]. Furthermore, the increase in RENs decreases the share of traditional generating units, resulting in insufficient operating reserves on the supply side [5]. As a consequence, the power fluctuations cannot be accommodated, and the safe operation of power systems is threatened [6]. Demand response (DR) can respond to the power system's operating reserve requirements by adjusting the power consumption of flexible loads on the demand side [7], thereby effectively mitigating power fluctuations and maintaining a better balance between supply and demand in power systems [8]. Therefore, in the foreseeable future, DR will become ever more vital in modern power systems [9].

Generally, the control methods of DR can be categorized into two types, namely centralized control and distributed control [10]. In centralized control, a control center is involved with overall decision-making authority, which gathers information on all flexible loads in DR, and dispatches the control commands directly for all these flexible loads [11]. However, centralized control presents a high requirement for communication facilities between the control center and the numerous flexible loads [12]. Distributed control is a promising solution to address this problem. This is because, in distributed control, only a few flexible loads need to receive signals from the control center, while most flexible loads can make autonomous decisions based on local feedback and neighboring information [13]. As a result, the communication burden between the control center and flexible loads can be released [14]. In addition, distributed control offers other advantages, such as scalability and flexibility, making it more suitable for controlling flexible loads in DR [15]. With the increasing application of advanced communication, information, and control technologies, DR has evolved into a cyber-physical system (CPS) with significantly improved performance [16].

However, deep cyber-physical coupling also introduces vulnerabilities to the control system of DR, i.e., potential cyber-attack threats. Cyber-attacks can disrupt the functionality of DR and render it incapable of meeting the power system's DR requirements. Actually, the threat of cyber-attacks on the control system of CPS is already an urgent problem with severe consequences [17]. For example, in December 2015, due

Manuscript received 22 September 2023; revised 1 February 2024; accepted 8 March 2024. Date of publication 25 March 2024; date of current version 29 October 2024. This work was supported by The Science and Technology Development Fund, Macau SAR under Grant 0003/2020/AKP, Grant SKL-IOTSC(UM)-2021-2023, and Grant FDCT/0022/2020/A1. Paper no. TPWRS-01503-2023. *(Corresponding author: Keng-Weng Lao.)*

Shaohua Yang, Keng-Weng Lao, and Hongxun Hui are with the State Key Laboratory of Internet of Things for Smart City and Department of Electrical and Computer Engineering, University of Macau, Macao 999078, China (e-mail: yc17436@um.edu.mo; johnnylao@um.edu.mo; hongxunhui@um.edu.mo).

Yulin Chen is with the Hainan Institute of Zhejiang University, Sanya 572025, China (e-mail: chenyl2017@zju.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2024.3381231.

Digital Object Identifier 10.1109/TPWRS.2024.3381231

to remote cyber-attacks on the supervisory control and data acquisition (SCADA) system, the Ukrainian power system suffered a blackout lasting several hours, affecting approximately 225,000 customers [18]. Furthermore, in March 2019, a devastating cyber-attack on industrial control systems plunged more than 20 states in Venezuela, including the capital city of Caracas, into darkness for over 24 hours [19]. This widespread power outage led to serious consequences, such as national traffic paralysis, the disruption of hospital operations, and the breakdown of communication lines [20]. The facts show that cyber-attacks are a severe threat to the control system in CPS and are becoming a necessary concern nowadays [21].

Likewise, the DR's control system, a typical CPS, is also subject to cyber-attack threats. Potential cyber-attacks that compromise system security include deception cyber-attacks (DCAs), denial-of-service attacks, delay attacks, replay attacks, and so on [22]. Among these, the DCA is regarded as the most typical attack form that threatens the control system, since it can destroy the control performance via tampered deceptive control information [23]. In order to deal with the adverse impact of attacks on load control on the demand side, some efforts have been made from the perspective of attack detection and localization [24], [25], [26]. For example, a two-dimensional convolutional neural network-based approach is introduced to detect and localize attacks on demand-side loads using data monitored by phasor measurement units (PMUs) in the power grid [24]. A protection scheme is designed against attacks on demand-side loads to address attack sensor location [25]. Furthermore, based on the attack detection and localization, a cyber-resilient economic dispatch framework is presented to mitigate the impact of attacks on demand-side loads while minimizing the operational cost by global optimization [26]. However, these existing efforts are based on a centralized framework, and assume that the control center can monitor the data of loads, which is not applicable to demand response (DR) with distributed control [27]. This is because in distributed control, the control center usually communicates with only a few flexible loads, and the information of most flexible loads is unknown [28]. As a result, there is not enough global information for state estimation, and it is hard to achieve attack detection and localization for DR with distributed control.

To address this cyber-security problem in DR with distributed control, Yang et al. [29] propose a resilient distributed control that operates without relying on detection. This innovative control method is the current state-of-the-art in DR distributed control under DCAs, which, for the first time, directly defends against static DCAs in distributed DR from the control perspective. However, beyond static DCAs, there exist various patterns of DCAs, such as scaling attacks, linear attacks, impulsive attacks, and so on [30]. When considering arbitrary attack patterns in distributed control of DR, significant research gaps remain to be filled. For example, diverse attack patterns can cause distinct impacts on DR. Therefore, revealing the specific impact of each attack pattern on DR remains to be resolved. Moreover, developing a secure control method for DR that can effectively counteract DCAs with arbitrary patterns has yet to be addressed.

To this end, this paper develops a novel secure distributed control for DR in the presence of DCAs with arbitrary patterns. The salient contributions of this work can be summarized below:

- We analyze attack impact on DR considering various patterns of DCAs, and reveal that different attack patterns are associated with distinct impacts. In particular, linear attacks, scaling attacks, and impulsive attacks result in power deviation, delayed response, and power fluctuation, respectively. Building upon this discovery, we formulate the cyber-security problem of DR, specifically addressing arbitrary attack patterns.
- We develop a novel anti-attack secure distributed (ASD) control for multi-load-based DR, ensuring resilience against DCAs on control. This ASD control method does not depend on detection and, for the first time, enables distributed DR to comprehensively counteract DCAs with arbitrary patterns.
- We provide rigorous proof based on Lyapunov theorem to demonstrate that, with the support of ASD control, power regulation in DR can always achieve the globally uniformly ultimately bounded (GUUB) convergence, even under arbitrary DCAs. That is, the developed ASD control can ensure that DR satisfies the power system's requirements, despite arbitrary DCAs.

II. Modelling of Cyber-Physical DR Community

In this section, a cyber-physical DR community is modeled by taking the heating, ventilation, and air conditioning (HVAC) as a flexible load example. First, the general framework of this HVAC-based DR community is given to show the DR process for offering operating reserves. Then, a detailed thermal and electrical model is presented for HVAC. Finally, distributed control is designed for the DR community, so that dispersed HVACs can work cooperatively to offer operating reserve services to the power system.

A. Overall Framework of Cyber-Physical DR Community

The overall framework of cyber-physical DR community with HVACs can be illustrated in Fig. 1. The focus of this paper is on HVACs because of the advantages of high regulation potential and low regulation influence (due to thermal inertia) [15]. It is worth noting that the cyber-physical DR community framework is also applicable to other flexible resources.

As shown in Fig. 1, during the DR process, the power system side needs to continuously monitor the system's operational status and determine the DR requirement. When facing power balance challenges, the power system operator will issue dispatch signals to the aggregator requesting operating reserves. The aggregator is a provider of DR services, who has the responsibility for responding to the power system operator. Upon receiving a dispatch signal, the aggregator will actively meet the requirements of the power system by adjusting the power consumption of HVACs. Typically, to offer considerable DR services, the aggregator has to aggregate a sufficient number of flexible loads. For this purpose, the dispersed HVAC-based

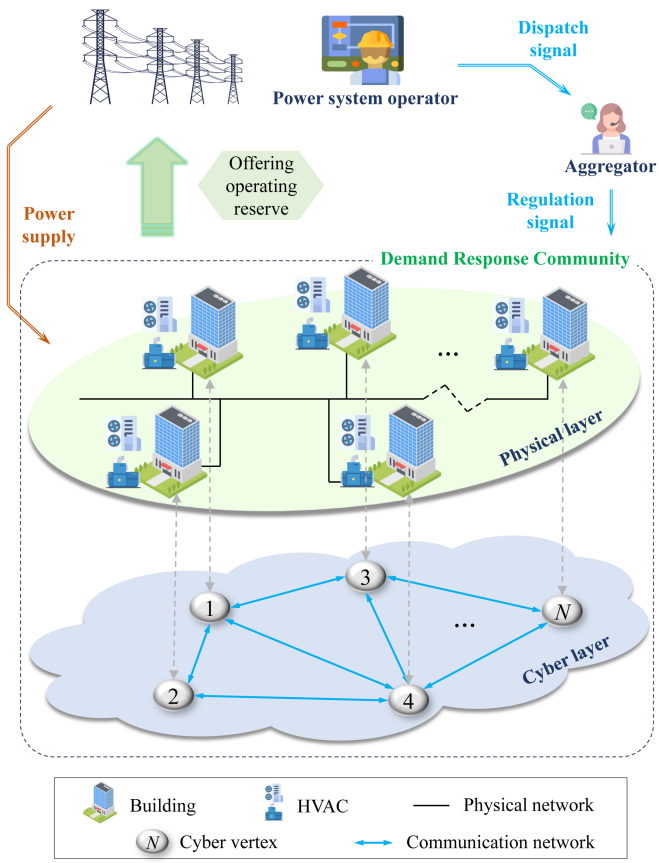

Fig. 1. Illustration of cyber-physical DR community to offer operating reserve services to the power system.

buildings are aggregated into a DR community. To control these decentralized HVACs in the DR community, distributed control is employed, which can avoid the burden of direct communication between the aggregator and individual HVACs, and can offer advantages such as plug-and-play capability. Based on distributed control, local HVAC individuals can communicate with each other to exchange information in cyber layer, so as to cooperatively accomplish the global objective toward the total power adjustment of the DR community in physical layer. In this manner, the required operating reserves for the power system can be satisfied, which means the DR service is accomplished.

B. Thermal and Electrical Model of HVAC

The operating power of HVAC is related to the building's thermal characteristics. Therefore, thermal modeling of the building is necessary for studying HVAC involved in DR services, which contains two crucial components, i.e., heat gain and heat loss, as follows [31]:

$$c_{\mathrm{a}}\rho_{\mathrm{a}}V\frac{dT_i(t)}{dt}=\underbrace{Q_{\mathrm{Gain}}(t)}_{\text{Heat gain}}-\underbrace{Q_{\mathrm{Loss}}(t)}_{\text{Heat loss}},\ \forall i\in\mathcal{I},\ \forall t\in\mathcal{T},\quad(1)$$

where c_a and ρ_a denote air heat capacity and air density, respectively; V is the building's volume; $T_i(t)$ is the indoor temperature of building i at time t; Q_{Gain} and Q_{Loss} denote the heat gain and heat loss, respectively; $\mathcal{I}$ denotes the set of HVACs in DR; and $\mathcal{T}$ is the set of time slots. In cooling mode, the heat gain generally comes from air exchange and heat transfer from outside, which can be expressed as follows:

$$Q_{\mathrm{Gain}}(t)=\underbrace{n_{\mathrm{ae}}c_{\mathrm{a}}\rho_{\mathrm{a}}V(T_{\mathrm{at}}-T_i(t))}_{\text{Air exchange}}+\underbrace{\Upsilon_{\mathrm{ht}}A_{\mathrm{s}}(T_{\mathrm{at}}-T_i(t))}_{\text{Heat transfer}},$$
$$\forall i\in\mathcal{I},\ \forall t\in\mathcal{T},\quad(2)$$

where n_{ae} and Υ_{ht} denote the time of air exchanges and the heat transfer coefficient, respectively; T_{at} denotes the ambient temperature; A_{s} denotes the envelope's surface area. Likewise, the heat loss comes from the cooling process of the HVAC, as follows:

$$Q_{\mathrm{Loss}}(t)=Q_{\mathrm{HVAC}}(t)=\kappa_Q\theta_i(t)P_{\mathrm{Rated}},\ \forall i\in\mathcal{I},\ \forall t\in\mathcal{T},\quad(3)$$

where κ_Q denotes the coefficient of performance (COP) of HVAC, which represents the relationship between cooling capacity and input electric power; P_{Rated} is the rated power of HVAC; and $\theta_i(t)$ denotes the power state of HVAC i at time t. Considering HVAC's participation in DR services, the power and thermal states are the two main concerns and can be defined as follows:

$$\begin{cases}\theta_i(t)=P_i(t)/P_{\mathrm{Rated}},\\ \psi_i(t)=(T_i(t)+\Delta T-T_{\mathrm{s}})/2\Delta T,\end{cases}\quad\forall i\in\mathcal{I},\ \forall t\in\mathcal{T},\quad(4)$$

where $P_i(t)$ is the operating power of HVAC i; $\psi_i(t)$ is the thermal state of HVAC i; ΔT and T_{s} denote the tolerated temperature change and the setting temperature, respectively. With the definition in (4), the concerned states can be normalized. Depending on the physical constraints, the power state has to be within $\theta_i\in[0,1]$, where the lower bound $\theta_i=0$ and the upper bound $\theta_i=1$ denote the HVAC i operates at zero power and P_{Rated}, respectively. In addition, considering the customer's thermal comfort requirements, the thermal state should also be within $\psi_i\in[0,1]$, where the lower bound $\psi_i=0$ and upper bound $\psi_i=1$ indicate that the indoor temperature T_i touches the tolerated cold limit $T_{\mathrm{Cold}}=T_{\mathrm{s}}-\Delta T$ and tolerated hot limit $T_{\mathrm{Hot}}=T_{\mathrm{s}}+\Delta T$, respectively.

In combination with (1)–(4), the dynamic model of thermal state can be derived as follows:

$$\frac{d\psi_i(t)}{dt}=-\frac{\kappa_Q\cdot P_{\mathrm{Rated}}}{2\Delta Tc_{\mathrm{a}}\rho_{\mathrm{a}}V}\theta_i(t)-\frac{(n_{\mathrm{ae}}c_{\mathrm{a}}\rho_{\mathrm{a}}V+\Upsilon_{\mathrm{ht}}A_{\mathrm{s}})}{c_{\mathrm{a}}\rho_{\mathrm{a}}V}\psi_i(t)$$
$$+\frac{(n_{\mathrm{ae}}c_{\mathrm{a}}\rho_{\mathrm{a}}V+\Upsilon_{\mathrm{ht}}A_{\mathrm{s}})(T_{\mathrm{at}}+\Delta T-T_{\mathrm{s}})}{2\Delta Tc_{\mathrm{a}}\rho_{\mathrm{a}}V},\ \forall i\in\mathcal{I},\ \forall t\in\mathcal{T}.\quad(5)$$

This dynamic model in (5) shows the relationship between the power and thermal states. This implies the thermal state ψ_i can be controlled by adjusting the power state θ_i of HVAC. Note that by making appropriate adjustments to the heat gain and heat loss components, the dynamic model of thermal state can still be extracted in modes other than the cooling mode.

For concision, $\mathbb{C}_{\mathrm{th}}=c_{\mathrm{a}}\rho_{\mathrm{a}}V$ and $\mathbb{G}_{\mathrm{th}}=n_{\mathrm{ae}}c_{\mathrm{a}}\rho_{\mathrm{a}}V+\Upsilon_{\mathrm{ht}}A_{\mathrm{s}}$ are utilized in the following text to denote the thermal capacitance

and thermal conductance coefficients, respectively. On this basis, the state-space model for the HVAC i can be derived:

$$\underbrace{\begin{bmatrix}\dot{\theta}_i(t)\\ \dot{\psi}_i(t)\end{bmatrix}}_{\dot{\boldsymbol{x}}_i} = \underbrace{\begin{bmatrix}0 & 0\\ -\frac{\kappa_Q\cdot P_{\text{Rated}}}{2\Delta T\mathbb{C}_{\text{th}}} & -\frac{\mathbb{G}_{\text{th}}}{\mathbb{C}_{\text{th}}}\end{bmatrix}}_{\mathscr{A}}\underbrace{\begin{bmatrix}\theta_i(t)\\ \psi_i(t)\end{bmatrix}}_{\boldsymbol{x}_i} + \underbrace{\begin{bmatrix}1\\ 0\end{bmatrix}}_{\mathscr{B}} v_i(t) + \underbrace{\begin{bmatrix}0\\ \frac{\mathbb{G}_{\text{th}}(T_{\text{at}}+\Delta T-T_{\text{s}})}{2\Delta T\mathbb{C}_{\text{th}}}\end{bmatrix}}_{\mathscr{C}}, \ \forall i\in\mathcal{I},\ \forall t\in\mathcal{T}. \tag{6}$$

where $\dot{[\,]}$ is the differentiation operator; $\boldsymbol{x}_i=[\theta_i,\psi_i]^{\text{T}}$ denotes the state variable vector, including power and thermal states described above; $\mathscr{A}$, $\mathscr{B}$, and $\mathscr{C}$ denote state transition matrix, input matrix, and supplementary matrix, respectively; v_i is the control input to be designed, as detailed below.

C. Distributed Cooperative Control for HVAC-Based DR

The DR community is an aggregated entity of several HVACs depicted by (6). Based on the distributed control design, the HVACs in the DR community can cooperatively offer DR services as required by the power system. When the customer-side HVACs enter into the DR process, the power system's power regulation requirements take top priority. Therefore, the power state θ_i is a pivotal adjustment variable for HVACs during the DR process, which can be selected for exchange information to realize distributed cooperative control. On this basis, $\forall i\in\mathcal{I}$, the HVAC's power state θ_i can be adjusted by considering the difference between the local and neighbors, along with the regulation signal sent from the aggregator, and the control protocol can be expressed:

$$\dot{\theta}_i = v_i = -k_\theta\sum_{j\in\mathcal{M}_{(i)}} a_{ij}(\theta_i-\theta_j)+b_i(\theta_i-\theta_{\text{reg}}), \tag{7}$$

where v_i denotes designed control input of HVAC i; k_θ is a coupling gain, which is a positive number; $\mathcal{M}_{(i)}$ stands for a set of all neighbors of HVAC i; a_{ij} denotes a entry in the adjacency matrix $\mathcal{A}$ of the communication topology, where $a_{ij}=1$ implies HVAC i and HVAC j exchange information with each other, and $a_{ij}=0$ otherwise; θ_{reg} is the regulation signal sent from the aggregator; b_i stands for a pinning gain, where $b_i=1$ indicates HVAC i can receive the regulation signal, and $b_i=0$ otherwise.

Considering all HVACs in the DR community, the corresponding matrix form of the control protocol (7) is expressed:

$$\boldsymbol{v} = -k_\theta(\mathcal{L}+\mathcal{B})\boldsymbol{\theta}+k_\theta\theta_{\text{reg}}\boldsymbol{b}, \tag{8}$$

where $\boldsymbol{v}=[v_1,v_2,\ldots,v_N]^{\text{T}}$ is the control vector; N represents the number of HVACs considered in the DR community; $\mathcal{L}=\mathcal{D}-\mathcal{A}$ is the Laplacian matrix, with $\mathcal{D}=\text{diag}\{d_i\}\subseteq\mathbb{R}^{N\times N}$ being the in-degree matrix of the communication topology and $d_i=\sum_{j\in\mathcal{M}_{(i)}}a_{ij}$; $\boldsymbol{\theta}=[\theta_1,\theta_2,\ldots,\theta_N]^{\text{T}}$; $\mathcal{B}=\text{diag}\{\boldsymbol{b}\}\subseteq\mathbb{R}^{N\times N}$ is the pinning matrix, with $\boldsymbol{b}=[b_1,b_2,\ldots,b_N]^{\text{T}}$.

Based on the distributed cooperative control above, the state-space model for the DR community can be obtained:

$$\underbrace{\begin{bmatrix}\dot{\boldsymbol{\theta}}\\ \dot{\boldsymbol{\psi}}\end{bmatrix}}_{\dot{\boldsymbol{x}}} = \underbrace{\begin{bmatrix}-k_\theta(\mathcal{L}+\mathcal{B}) & \boldsymbol{0}\\ -\frac{\kappa_Q\cdot P_{\text{Rated}}}{2\Delta T\mathbb{C}_{\text{th}}}\boldsymbol{I}_N & -\frac{\mathbb{G}_{\text{th}}}{\mathbb{C}_{\text{th}}}\boldsymbol{I}_N\end{bmatrix}}_{\mathscr{T}}\underbrace{\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\psi}\end{bmatrix}}_{\boldsymbol{x}} + \underbrace{\begin{bmatrix}k_\theta\theta_{\text{reg}}\mathcal{B} & \boldsymbol{0}\\ \boldsymbol{0} & \frac{\mathbb{G}_{\text{th}}(T_{\text{at}}+\Delta T-T_{\text{s}})}{2\Delta T\mathbb{C}_{\text{th}}}\boldsymbol{I}_N\end{bmatrix}}_{\mathscr{D}}\underbrace{\begin{bmatrix}\boldsymbol{1}_N\\ \boldsymbol{1}_N\end{bmatrix}}_{\mathbf{d}}, \tag{9}$$

where $\boldsymbol{x}\subseteq\mathbb{R}^{2N}$ is the vector of state variables; $\mathscr{T}$ and $\mathscr{D}$ are the state transition matrix and supplementary matrix of the DR community; $\boldsymbol{I}_N\subseteq\mathbb{R}^{N\times N}$ denotes the identity matrix; $\mathbf{d}\subseteq\mathbb{R}^{2N}$ is a 1-vector. From the state-space model in (9), it is known that in the DR community, the thermal state of HVAC can be controlled. Also, the HVAC's power state can be adjusted according to the aggregator's regulation signal, which implies the DR community can offer operating reserve services to power systems.

Remark 1: The proposed framework can be robust enough to handle communication interruptions of a_{ij} and b_i, if the communication topology of distributed control contains a spanning tree, and at least one HVAC can receive regulation signals from the aggregator.

III. Anti-Attack Secure Distributed Control for DR

In this part, we first formulate DR's cyber-security problem under DCAs with different patterns. After that, an anti-attack secure distributed (ASD) control is developed for DR to address this problem. Finally, it is rigorously proved by Lyapunov theorem that the GUUB convergence can be guaranteed even under DCA with arbitrary patterns.

A. DR's Cyber-Security Problem Due to Arbitrary DCAs

As appealed in the white paper 'Cybersecurity for Industrial Automation and Control Environments,' cyber-attacks have become an increasingly severe threat to industry control systems [32]. In practice, there are cyber-security risks in the communication between HVAC devices and controllers in DR [33]. The transmitted control signal can be intercepted and tampered with by hackers, and the DR task can fail due to the malicious DCA on the control system.

1) General Form of Different DCA Patterns: Typical DCA patterns can be classified into the following six categories, i.e., (1) shift attack, (2) scaling attack, (3) linear attack, (4) non-linear attack, (5) impulsive attack, and (6) interruption attack, and all of them are described in detail as below.

The shift attack can be illustrated as follows:

$$\widetilde{v_i}(t) = v_i(t)+\sum_{h\in\mathcal{H}}\phi_{ih},\ \forall i\in\mathcal{I},\ \forall t\in\mathcal{T}, \tag{10}$$

where $v_i(t)$ denotes the original control input of HVAC i; ϕ_{ih} represents the shift attack data h imposed on controller i, which can offset the original signal; $\mathcal{H}$ refers to the set encompassing all shift attack data; and $\widetilde{v_i}(t)$ denotes the corrupted control signal by malicious DCAs.

The scaling attack is another type of DCA, which can scale the original control input:

$$\widetilde{v}_i(t) = s_i \cdot v_i(t),\ \forall i \in \mathcal{I},\ \forall t \in \mathcal{T}, \tag{11}$$

where s_i represents the scaling parameter.

The linear attack's general form can be represented by combining the shift and scaling attacks described above [30]:

$$\widetilde{v}_i(t) = s_i \cdot v_i(t) + \sum_{h\in\mathcal{H}} \phi_{ih},\ \forall i \in \mathcal{I},\ \forall t \in \mathcal{T}. \tag{12}$$

The non-linear attack can be regarded as replacing the original information with a new non-linear signal [34], which can be modeled as follows:

$$\widetilde{v}_i(t) = f_i(t),\ \forall i \in \mathcal{I},\ \forall t \in \mathcal{T}, \tag{13}$$

where $f_i(t)$ is the replaced signal for the controller i, which can be an arbitrary pattern, e.g., cosine function.

Moreover, when the impulsive attack occurs, the control signal can be represented as follows [35]:

$$\widetilde{v}_i(t) = v_i(t) + \sum_{k=1}^{n} d_k v_i(t)\delta(t - t_k),\ \forall i \in \mathcal{I},\ \forall t \in \mathcal{T}, \tag{14}$$

where d_k denotes the destabilizing impulse parameter; $\delta(\cdot)$ denotes the Dirac impulse; and $\{t_k\}_{k=1}^{n}$ is a sequence of n impulses.

In addition, the interruption attack can be shown as:

$$\widetilde{v}_i(t) = v_i(t) + (-v_i(t)),\ \forall i \in \mathcal{I},\ \forall t \in \mathcal{T}, \tag{15}$$

where $-v_i(t)$ indicates the interruption attack signal launched by a hacker, which is opposite to the original signal. It is clear from the (15) that under interruption attack, $\widetilde{v}_i(t)$ is equal to 0, i.e., the control signal is interrupted.

It is worth noting that each attack pattern takes a different form and changes the transmitted data differently. The shift attack directly injects extra false data, and the scaling attack scales the original data up or down. The linear deception attack can be a combination of the shift and scaling attacks, while the non-linear attack can usually replace the original signal with a new non-linear signal directly. The impulsive attack can be thought of as a shift attack carried out in an impulsive manner. Moreover, the interruption attack can interrupt the transmitted signal. To consider different DCAs comprehensively, all of them can be reformulated into a general form:

$$\widetilde{v}_i(t) = v_i(t) + \xi_i,\ \forall i \in \mathcal{I},\ \forall t \in \mathcal{T}, \tag{16}$$

where ξ_i is the equivalent injection data of HVAC i, which can be regarded as the equivalent attack on control inputs by different attack patterns, as detailed in Table I. With this basis, all these attack patterns can be analyzed according to this general form with attack injection ξ_i.

2) DR's Cyber-Security Problem Formulation: As illustrated in Fig. 2, malicious hackers can inject exogenous signals to disturb the HVAC's local control input channel, and aim to destabilize the HVAC-based DR community by inserting these DCAs. On this basis, instead of (7), one has:

$$\dot{\theta}_i = \widetilde{v}_i = v_i + \mu_i \xi_i,\ \forall i \in \mathcal{I}, \tag{17}$$

TABLE I
GENERAL FORM OF DIFFERENT DCA PATTERNS AND THEIR EQUIVALENT EXPRESSIONS

General form of attack patterns $\widetilde{v}_i(t) = v_i(t) + \xi_i$	
Attack pattern	Equivalent expression of ξ_i
Shift attack	$\xi_i = \sum_{k\in\mathcal{K}} \phi_{ik}$
Scaling attack	$\xi_i = [s_i - 1] \cdot v_i(t)$
Linear attack	$\xi_i = [s_i - 1] \cdot v_i(t) + \sum_{k\in\mathcal{K}} \phi_{ik}$
Non-linear attack	$\xi_i = [f_i(t) - v_i(t)]$
Impulsive attack	$\xi_i = \sum_{k=1}^{n} d_k v_i(t)\delta(t - t_k)$
Interruption attack	$\xi_i = -v_i(t)$

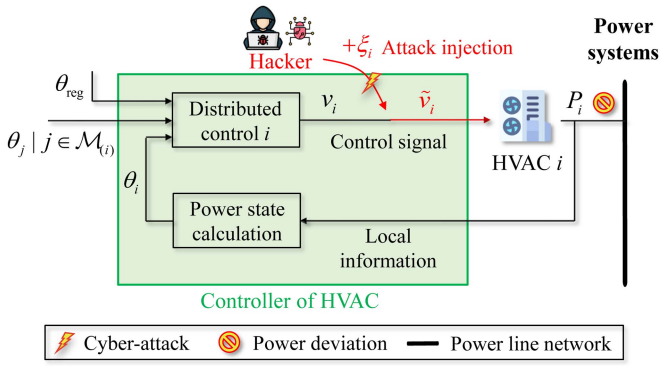

Fig. 2. Cyber-attack on the HVAC's controller.

where μ_i is the occurring parameter with a binary number; $\mu_i = 1$ implies that the cyber-attack occurred for HVAC i, otherwise, $\mu_i = 0$. Note that attack injection ξ_i can be arbitrary patterns. Considering the numerous HVACs in the DR community, the attacks launched by the hacker can also be characterized in vector form as $\boldsymbol{\xi} = [\mu_1\xi_1, \mu_2\xi_2, \ldots, \mu_N\xi_N]^{\mathrm{T}}$, which represents the attack vector.

In addition, the following assumption is held in this article.

Assumption 1: For all $i \in \mathcal{I}$, the attack injection ξ_i launched by the hacker is not infinite, but bounded.

This assumption is reasonable since the actuator is subject to physical constraints. In other words, there is a limitation to the actual impact of the attack data launched by hackers due to the practical physical constraints. In particular, the following boundary exists for the attack injection:

$$||\boldsymbol{\xi}|| \le \rho, \tag{18}$$

where $||\cdot||$ denotes the 2-norm; ρ is a positive constant, indicating the parameter limit for attack due to physical boundary.

To account for the impact of attack, the regulation error for the HVAC i's power state is defined as follows:

$$\epsilon_i = \theta_i - \theta_{\text{reg}},\ \forall i \in \mathcal{I}, \tag{19}$$

and its matrix form for the global regulation error of the overall HVACs in the DR community can be given:

$$\boldsymbol{\epsilon} = \boldsymbol{\theta} - \theta_{\text{reg}}\mathbf{1}_N, \tag{20}$$

where $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \ldots, \epsilon_N]^{\mathrm{T}}$.

Consider the power state dynamics under attack in (17) and the defined regulation error in (19). We can obtain the following closed-loop regulation error dynamics:

$$\dot{\boldsymbol{\epsilon}} = -k_\theta(\mathcal{L}+\mathcal{B})\boldsymbol{\epsilon} + \boldsymbol{\xi}. \tag{21}$$

Since attack injection $\boldsymbol{\xi}$ is an equivalent expression, $\boldsymbol{\xi}$ in each attack pattern will have different impacts. For example, when the HVAC's controller is hacked by a shift attack in (10), it can be derived that the regulation error for HVAC i is:

$$\epsilon_i = \sum_{k\in\mathcal{I}_{attack}} \left(\gamma_{ik}\sum_{h\in\mathcal{H}}\phi_{kh}\right),\ \forall i\in\mathcal{I}, \tag{22}$$

where $\mathcal{I}_{attack}$ is the set of attacked HVACs; γ_{ik} represents the entries of the attack influence matrix $\boldsymbol{\Gamma} = [k_\theta(\mathcal{L}+\mathcal{B})]^{-1}$; $\sum_{h\in\mathcal{H}}\phi_{kh}$ is the shift attack data for HVAC k. This implies that a power deviation in HVAC i is associated with an attack on HVAC k, and a shift attack can result in steady-state errors in the power regulation of all HVACs. Whereas under scaling attack in (11), the regulation error is:

$$\epsilon_i = 0,\ \forall i\in\mathcal{I}, \tag{23}$$

which means the scaling attack will not affect the final steady-state result of the power regulation. In fact, it will only affect the response speed. However, when the shift attack pattern and the scaling attack pattern are combined, i.e., under a linear attack in (12), the regulation error can be derived as:

$$\epsilon_i = \sum_{k\in\mathcal{I}_{attack}} \left(\frac{1}{s_k}\gamma_{ik}\sum_{h\in\mathcal{H}}\phi_{kh}\right),\ \forall i\in\mathcal{I}, \tag{24}$$

where s_k is the scaling parameter for HVAC k. It can be seen that the error can be amplified by adjusting the scaling parameter. Moreover, in the case of the non-linear and impulsive attacks, the regulation error will not converge to a steady state but continuously fluctuate with dynamic attack injection.

Remark 2: By analyzing, different DCA patterns can cause various adverse consequences on DR. Under different attack patterns, DR can exhibit a sluggish response to the power system, failing to meet the required operating reserve services, or even introducing harmful power fluctuations to the power system. These adverse impacts can jeopardize the safe operation of the power system and must be addressed.

Remark 3: Some smart hackers can adapt their attack models based on specific information to amplify the attack effect, or use advanced algorithms to evade detection. However, there are also many hackers who indiscriminately attack without considering the attack effect or detection. Both types of attack behaviors are worth investigating. Therefore, arbitrary patterns of the DCA should be defended.

Remark 4: In distributed control, there is not enough global information for state estimation [27]. For this reason, detecting and localizing the attacked HVACs is also hardly performed in HVAC-based DR with distributed control. Therefore, a more direct solution without relying on detection has to be taken to address the cyber-security problem, i.e., developing an advanced control strategy.

Therefore, the problem can be formulated as follows:

Problem 1: It is essential to develop an advanced secure distributed control strategy for DR to counteract against DCAs with arbitrary patterns.

B. Anti-Attack Secure Distributed Control Development

In order to defend against arbitrary DCAs, an anti-attack secure distributed (ASD) control is developed to safeguard the DR community in power systems:

$$\phi_i = -k_\theta\sum_{j\in\mathcal{M}_{(i)}} a_{ij}(\theta_i-\theta_j) + b_i(\theta_i-\theta_{\text{reg}}) + \zeta_i, \tag{25}$$

$$\zeta_i = \frac{\rho^2 v_i}{\omega},\ \forall i\in\mathcal{I}, \tag{26}$$

where ϕ_i is control input of the ASD control for HVAC i; ζ_i is the adaptive compensation term; and ω is a positive constant.

Theorem 1: Consider arbitrary DCAs in (17). Let the ASD control protocol consist of (25) and (26). Then, the global regulation error of DR community $\boldsymbol{\epsilon}$ described in (20) can achieve GUUB convergence. That is, the DR's cyber-security problem is solved.

To assess the convergence results of the developed ASD control, the following stability result is defined.

Definition 1: [36]: $x(t)\in\mathbb{R}$ is GUUB with ultimate bound b, if there exist positive constants b, independent of $t_0\geq 0$, and for arbitrary a, there is $t_x = t_x(a,b) > 0$, independent of t_0, such that

$$||x(t_0)||\leq a \Rightarrow ||x(t)||\leq b,\ \forall t\geq t_0+t_x. \tag{27}$$

Now, we prove the Theorem 1. It is worth emphasizing that the following theoretical derivation is original.

Proof: Oriented to numerous HVACs, the ASD control protocol needs to be expressed in matrix form:

$$\boldsymbol{\phi} = -k_\theta(\mathcal{L}+\mathcal{B})\boldsymbol{\theta} + k_\theta\theta_{\text{reg}}\boldsymbol{b} + \boldsymbol{\zeta}, \tag{28}$$

where $\boldsymbol{\zeta} = [\zeta_1,\zeta_2,\ldots,\zeta_N]^{\text{T}}$ is the compensation term in matrix form.

The derivative of (8) yields:

$$\dot{\boldsymbol{v}} = -k_\theta(\mathcal{L}+\mathcal{B})\dot{\boldsymbol{\theta}}. \tag{29}$$

Considering an arbitrary DCA vector $\boldsymbol{\xi}$ and the proposed ASD control, the power state dynamics can be obtained as follows:

$$\dot{\boldsymbol{\theta}} = \boldsymbol{v} + \boldsymbol{\zeta} + \boldsymbol{\xi}. \tag{30}$$

Then, consider a Lyapunov function candidate as below:

$$E = \frac{1}{2}\boldsymbol{v}^T\boldsymbol{v}, \tag{31}$$

and combining it with (29) and (30) yields the time-derivative of this Lyapunov function candidate:

$$\begin{aligned}\dot{E} &= \boldsymbol{v}^T\dot{\boldsymbol{v}}\\ &= \boldsymbol{v}^T[-k_\theta(\mathcal{L}+\mathcal{B})\dot{\boldsymbol{\theta}}]\\ &= \boldsymbol{v}^T[-k_\theta(\mathcal{L}+\mathcal{B})(\boldsymbol{v}+\boldsymbol{\zeta}+\boldsymbol{\xi})]\end{aligned}$$

$$= \boldsymbol{v}^T[-k_\theta(\mathcal{L}+\mathcal{B})]\boldsymbol{v} + \boldsymbol{v}^T[-k_\theta(\mathcal{L}+\mathcal{B})](\boldsymbol{\zeta}+\boldsymbol{\xi}). \quad (32)$$

Combining (26) and (32), and applying the method of enlarging and reducing yields:

$$\begin{aligned}\dot{E} &= -\boldsymbol{v}^T[k_\theta(\mathcal{L}+\mathcal{B})]\boldsymbol{v} - \boldsymbol{v}^T[k_\theta(\mathcal{L}+\mathcal{B})]\left(\frac{\rho^2}{\omega}\boldsymbol{v}+\boldsymbol{\xi}\right)\\ &= -\boldsymbol{v}^T[k_\theta(\mathcal{L}+\mathcal{B})]\boldsymbol{v} - \frac{\rho^2}{\omega}\boldsymbol{v}^T[k_\theta(\mathcal{L}+\mathcal{B})]\boldsymbol{v}\\ &\quad - \boldsymbol{v}^T[k_\theta(\mathcal{L}+\mathcal{B})]\boldsymbol{\xi}\\ &\le -\boldsymbol{v}^T[k_\theta(\mathcal{L}+\mathcal{B})]\boldsymbol{v} - \frac{\rho^2}{\omega}\boldsymbol{v}^T[k_\theta(\mathcal{L}+\mathcal{B})]\boldsymbol{v}\\ &\quad + ||\boldsymbol{v}|| \cdot ||k_\theta(\mathcal{L}+\mathcal{B})\cdot\boldsymbol{\xi}||. \end{aligned} \quad (33)$$

Recalling Courant–Fischer–Weyl theorem [37] and noting that $k_\theta(\mathcal{L}+\mathcal{B})$ is positive-definite, it then follows that:

$$\sigma_{\min} \le \frac{||k_\theta(\mathcal{L}+\mathcal{B})\cdot\boldsymbol{\beta}||}{||\boldsymbol{\beta}||} \le \sigma_{\max}, \quad (34)$$

where $\boldsymbol{\beta}$ denotes a column vector; $\sigma_{\max}$ and $\sigma_{\min}$ are the maximum and minimum singulars of the matrix $k_\theta(\mathcal{L}+\mathcal{B})$, respectively.

Then, recalling Eigendecomposition of a matrix and combining (33) and (34), one can obtain:

$$\begin{aligned}\dot{E} &\le -\sigma_{\min}\boldsymbol{v}^T\boldsymbol{v} - \frac{\rho^2}{\omega}\sigma_{\min}\boldsymbol{v}^T\boldsymbol{v} + \sigma_{\max}||\boldsymbol{v}||\cdot||\boldsymbol{\xi}||\\ &\le -\sigma_{\min}\boldsymbol{v}^T\boldsymbol{v} + \sigma_{\max}\rho||\boldsymbol{v}||\left(1 - ||\boldsymbol{v}||\frac{\rho\sigma_{\min}}{\omega\sigma_{\max}}\right). \end{aligned} \quad (35)$$

Choosing $||\boldsymbol{v}|| > \frac{\omega\sigma_{\max}}{\rho\sigma_{\min}}$ yields:

$$\dot{E} \le -\sigma_{\min}||\boldsymbol{v}||^2 + 0. \quad (36)$$

In this case, the derivative of Lyapunov function candidate $\dot{E} < 0$, i.e., the system described by Lyapunov function in (31) is negative-definite. Hence, this system is globally uniformly stable (UGS). That is, conditional stability can be achieved.

When $||\boldsymbol{v}|| \le \frac{\omega\sigma_{\max}}{\rho\sigma_{\min}}$, it yields:

$$0 \le \left(1 - ||\boldsymbol{v}||\frac{\rho\sigma_{\min}}{\omega\sigma_{\max}}\right) \le 1. \quad (37)$$

Combining (31), (35) and (37) yields:

$$\begin{aligned}\dot{E} &\le -\sigma_{\min}\boldsymbol{v}^T\boldsymbol{v} + \sigma_{\max}\rho||\boldsymbol{v}||\\ &\le -2\sigma_{\min}E + \frac{\omega\sigma_{\max}^2}{\sigma_{\min}}. \end{aligned} \quad (38)$$

Introduce a positive function $\lambda(t) > 0$ to satisfy:

$$\dot{E} + \lambda(t) = -2\sigma_{\min}E + \frac{\omega\sigma_{\max}^2}{\sigma_{\min}}. \quad (39)$$

We can find that (39) can be reformulated into a non-homogeneous differential equation:

$$\dot{E} + 2\sigma_{\min}E = -\lambda(t) + \frac{\omega\sigma_{\max}^2}{\sigma_{\min}}. \quad (40)$$

The solution of this differential equation is as follows:

$$\begin{aligned}E(t) &= E(0)\cdot e^{-2\sigma_{\min}t} - e^{-2\sigma_{\min}t}\cdot\int_0^t e^{2\sigma_{\min}\tau}\lambda(t)\mathrm{d}\tau\\ &\quad + \frac{\omega\sigma_{\max}^2}{\sigma_{\min}}e^{-2\sigma_{\min}t}\cdot\int_0^t e^{2\sigma_{\min}\tau}\mathrm{d}\tau. \end{aligned} \quad (41)$$

Since $e^{-2\sigma_{\min}t}$, $e^{2\sigma_{\min}\tau}$, and $\lambda(t)$ are all positive, the second term of the right side of (41) is less than zero. Then, we have:

$$\begin{aligned}E(t) &\le E(0)\cdot e^{-2\sigma_{\min}t} + \frac{\omega\sigma_{\max}^2}{\sigma_{\min}}e^{-2\sigma_{\min}t}\cdot\int_0^t e^{2\sigma_{\min}\tau}\mathrm{d}\tau\\ &= E(0)\cdot e^{-2\sigma_{\min}t} + \frac{\omega\sigma_{\max}^2}{2\sigma_{\min}^2}e^{-2\sigma_{\min}t}\cdot e^{2\sigma_{\min}\tau}|_0^t\\ &= E(0)\cdot e^{-2\sigma_{\min}t} + \frac{\omega\sigma_{\max}^2}{2\sigma_{\min}^2}(1 - e^{-2\sigma_{\min}t}). \end{aligned} \quad (42)$$

Combining (31) and (42) yields:

$$\begin{aligned}E(t) &= \frac{1}{2}\boldsymbol{v}(t)^T\boldsymbol{v}(t) = \frac{1}{2}||\boldsymbol{v}(t)||^2\\ &\le \frac{1}{2}\boldsymbol{v}(0)^T\boldsymbol{v}(0)\cdot e^{-2\sigma_{\min}t} + \frac{\omega\sigma_{\max}^2}{2\sigma_{\min}^2}(1 - e^{-2\sigma_{\min}t}). \end{aligned} \quad (43)$$

Note that $e^{-2\sigma_{\min}t} \to 0$ as $t \to \infty$, then we have:

$$\lim_{t\to\infty}||\boldsymbol{v}(t)|| \le \frac{\sigma_{\max}\sqrt{\omega}}{\sigma_{\min}}. \quad (44)$$

Hence, the variable $\boldsymbol{v}$ is GUUB. Combining (8), (20), and (44) yields:

$$\begin{aligned}\lim_{t\to\infty}||\boldsymbol{\epsilon}(t)|| &= \lim_{t\to\infty}||[k_\theta(\mathcal{L}+\mathcal{B})]^{-1}\cdot\boldsymbol{v}(t)||\\ &\le \lim_{t\to\infty}||[k_\theta(\mathcal{L}+\mathcal{B})]^{-1}||\cdot||\boldsymbol{v}(t)||\\ &\le \frac{\sigma_{\max}\sqrt{\omega}}{\sigma_{\min}}\cdot||[k_\theta(\mathcal{L}+\mathcal{B})]^{-1}||. \end{aligned} \quad (45)$$

Therefore, $\exists\ \tau_s > 0$, such that $\forall t > \tau_s$, $||\boldsymbol{\epsilon}(t)||$ is bounded by $\frac{\sigma_{\max}\sqrt{\omega}}{\sigma_{\min}}\cdot||[k_\theta(\mathcal{L}+\mathcal{B})]^{-1}||$. That is, the DR's global regulation error $\boldsymbol{\epsilon}$ described in (20) can achieve GUUB convergence, despite arbitrary DCAs.

The proof is complete. ■

Remark 5: With the proposed ASD control, for each HVAC $i \in \mathcal{I}$ in the DR community, the power state θ_i can converge to a tiny neighborhood of the desired regulation value θ_{reg}. That is, the adverse impacts due to arbitrary DCAs can be counteracted, and the DR's total power can be controlled by the aggregator according to the power system's requirements. In other words, with the proposed control, DR tasks can still be realized for power systems even under arbitrary DCAs.

Remark 6: It is worth noting that the defense mechanism of the proposed solution does not rely on anomaly identification or detection. Instead, it is based on a control approach to counter against cyber-attacks directly. For this reason, the proposed solution only involves modifying the HVAC's control strategy during DR, without additional monitoring devices for state estimation and data detection. Therefore, the proposed ASD control is a feasible implementation solution.

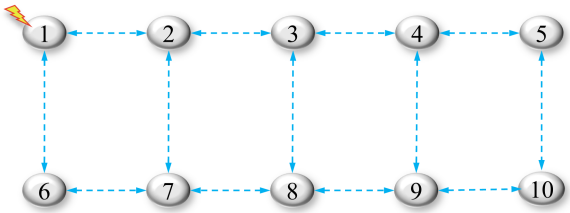

Fig. 3. Communication topology of HVACs in the DR community.

TABLE II
TYPICAL PARAMETERS FOR THE HVAC-BASED DR COMMUNITY

Symbols	Parameters	Values	Units
c_{a}	Air heat capacity	1.005	kJ/(kg·°C)
ρ_{a}	Air density	1.205	kg/m^3
H	Height of building	60	m
S	Floor area	5000	m^2
n_{ae}	Air exchange times	0.5	1/h
Υ_{ht}	Heat transfer coefficient	7.69	W/(m^2·°C)
P_{Rated}	Rated power	331	kW
κ_Q	COP	6.9	–
P_{intl}	Initial power	$\mathcal{U}(215{,}315)$[1]	kW
T_{at}	Ambient temperature	31	°C
T_{s}	Set temperature	25	°C
ΔT	Tolerable temperature change	± 2	°C
T_{intl}	Initial indoor temperature	$\mathcal{U}(23{,}27)$	°C

[1] $\mathcal{U}$ denotes uniform distributions.

In addition, the proposed ASD control can be scalable to other distributed control application scenarios, such as large-scale power systems with diverse power generation sources, to defend against potential cyber-security risks.

IV. Case Study and Verification

A. Test System

In this section, the efficacy of the proposed ASD control is verified in a DR community with 10 buildings. Each building is equipped with an inverter-based HVAC, i.e., a central air conditioner model LSBLX650SVE, which can be applied to regulate the temperature of a whole building [38]. The communication topology of these HVACs is shown in Fig. 3, where each HVAC can communicate locally with its neighbors. The parameters of the HVAC, such as the COP and rated power, are from realistic testing according to ARI550/590-2003 standard operating conditions [38], [39]. In addition, the ambient temperature parameters are realistic testing data as of August 28th, 2023, in Macau [40]. The HVAC parameters, corresponding building parameters, and ambient parameters [41] are presented in detail in Table II.

The test follows a process outlined below: At 14:00, the aggregator receives a DR dispatch signal from the power system operator, containing information on the required regulation capacity (750 kW) and duration of the DR task (1 h). As the manager of the DR community, the aggregator is responsible for offering operating reserves to the power system. Therefore, by controlling the power state of each HVAC, the aggregator adjusts the DR community's total power with the expectation of meeting the power system's DR requirements. Also, attacks are launched at the beginning of the test (14:00) that can disrupt the 1st HVAC's local control input channel. In addition, the normal control presented in equations (7) to (8) is employed at first, and the secure control, i.e., the proposed ASD control, is activated at the halfway point of the test (14:30) to defend against attacks until the end of the test. The test lasts until 15:00 for a total test time of 1 h.

There are five scenarios considering different DCA patterns: (S-1) shift attack pattern, (S-2) scaling attack pattern, (S-3) linear attack pattern, (S-4) non-linear attack pattern, and (S-5) impulsive attack pattern.

B. Scenario 1: Performance With Shift Attack Pattern

In scenario 1, we consider the shift attack pattern described in (10), where the attack is injected at the local control input of the 1st HVAC with a shift attack data of 0.1. The performance of the HVAC-based DR under attack and the efficacy of the proposed ASD control can be shown in Fig. 4.

As shown in Fig. 4(a), under the shift attack, all HVACs in DR have severe power state deviations, and the hacker can affect multiple HVACs at the same time. This is because there exist interactions between the cyber and physical layers, as well as interactions between the local HVAC and the neighboring HVACs in the DR community. Due to these deep couplings, the attack impact can be propagated and amplified. As a consequence, the DR community's total power also deviates significantly from the target total power required by the power system's DR task (response gap is about 264.8 kW),[1] as shown in Fig. 4(b). Observing over time, when the test is halfway conducted, the deviations are eliminated in a short period of time. The response speed is approximately 22.0 s and the consensus is achieved in the power state of all HVACs. This is because, at 1800 s, the proposed ASD control is activated. The test results imply the proposed ASD control can effectively counteract the adverse impact caused by shift attacks.

Define completion rate (CR) as an indicator for DR tasks:

$$\chi(t) = \frac{TP_{\mathrm{intl}} - TP(t)}{P_{\mathrm{reg}}}\%, \tag{46}$$

where $\chi(t)$ is the CR's value at time t; TP_{intl} and $TP(t)$ are the total power of DR community at the initial time and time t, respectively; P_{reg} is the regulation capacity required by power systems. Moreover, the average CR denotes the average completion rate during the test period.

From Fig. 4(c), it can be seen that under the shift attack, the average CR can only achieve about 64.25%, which implies the DR can no longer offer operating reserve services to the power system as required. However, with the proposed ASD control, the CR is recovered rapidly, and the average CR can be improved to about 99.84%, even under attack.

[1] In order to offer 750 kW of regulation capacity to the power system, the DR community's target total power was supposed to be reduced from the initial value of 2694.3 kW to 1944.3 kW. However, due to the cyber-attack, the total power can only be regulated to approximately 2209.1 kW.

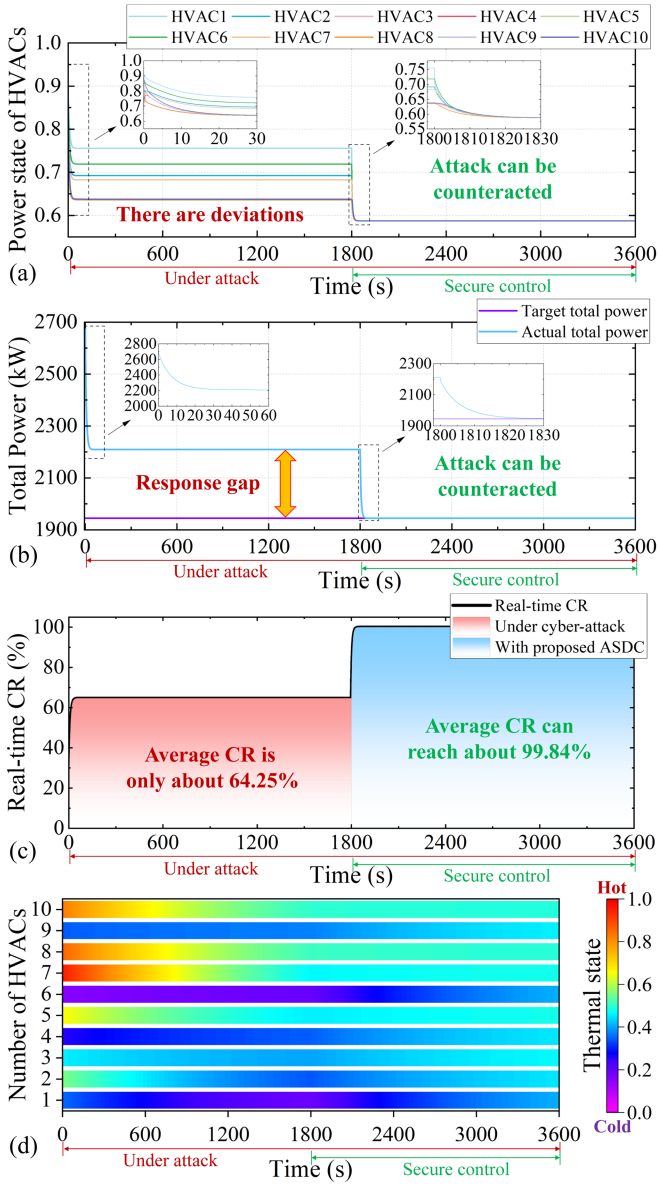

Fig. 4. Performance of the HVAC-based DR under shift attack and efficacy of the proposed ASD control: (a) power state of each HVAC; (b) DR community's total power; (c) real-time CR of DR; and (d) thermal state.

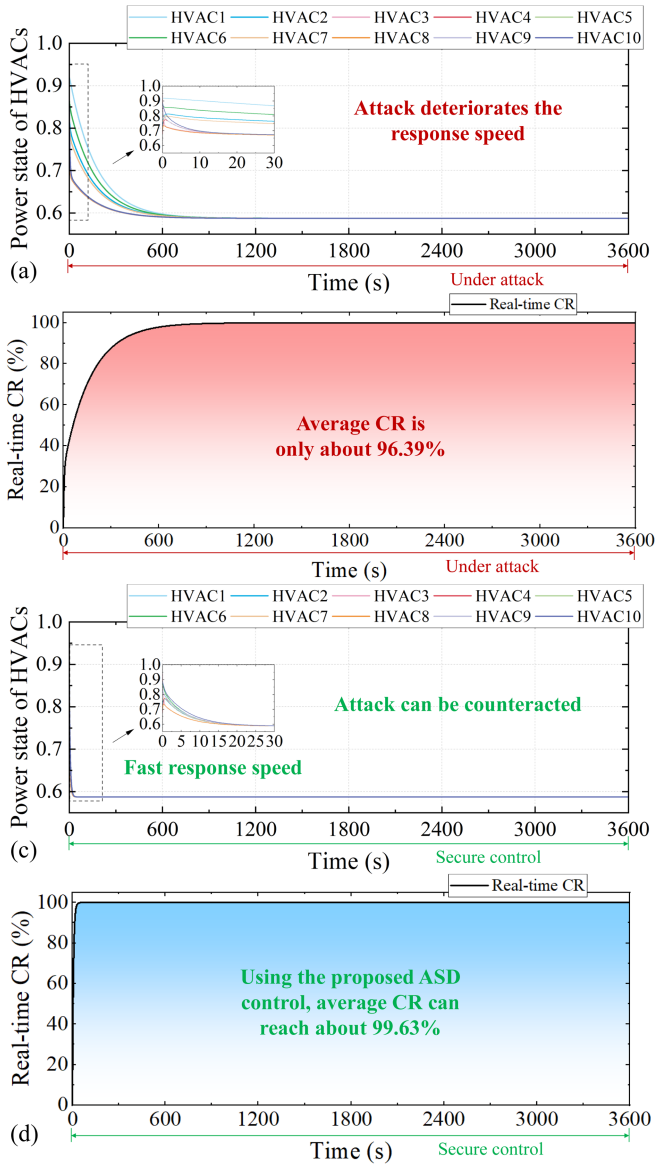

Fig. 5. Performance of the HVAC-based DR under scaling attack and efficacy of the proposed ASD control: (a) power state of each HVAC; (b) DR community's total power; (c) real-time CR of DR; and (d) thermal state.

In addition, as shown in Fig. 4(d), under the attack, the thermal states of the 1-th and 6-th HVACs tend towards the tolerated cold limit gradually, which implies the indoor temperature is affected by the attack non-negligibly. However, with the proposed ASD control, the thermal state of each HVAC can be gradually restored to the comfortable temperature region.

This scenario shows that the proposed ASD control can help DRs to have remarkable resilience against shift attacks.

C. Scenario 2: Performance With Scaling Attack Pattern

In scenario 2, we consider the scaling attack pattern described in (11), where the attack can scale the HVAC's local control inputs by a ratio of 0.01. The performance of HVAC-based DR is shown in Fig. 5.

From Fig. 5(a) and (b), it can be seen that under the scaling attack, the power states of each HVAC can still achieve consensus convergence, and the DR service can still be completed. This implies the scaling attack does not affect the final steady-state results. However, under the influence of scaling attack, the convergence and response speed become sluggish, taking about 731.7 s. It is noteworthy that power systems typically specify that DR services should be available within a predetermined timeframe, e.g., a response time within 5 minutes [42]. This indicates that a scaling attack can prevent the DR community from accomplishing DR services within the scheduled time. Meanwhile, the average CR is only about 96.39% due to the deteriorated response speed, which indicates deficiencies in the quality of DR services.

The performance with the proposed ASD control is shown in Fig. 5(c) and (d). It can be seen that with the proposed control, the HVAC's power state converges quickly, even under scaling attacks. As a result, the DR service also responds quickly, taking approximately 29.1 s, which means a significant prompt in response time. As well, the average CR can reach about 99.63%, which implies the DR service quality is effectively improved. Hence, the proposed ASD control can defend against scaling attacks and protect the DR community well.

D. Scenario 3: Performance With Linear Attack Pattern

In scenario 3, we consider the linear attack pattern described in (12), i.e., the simultaneous presence of scaling attack pattern and shift attack pattern (data is 0.1 again, as in scenario 1). The performance is shown in Fig. 6.

As shown in Fig. 6(a), the linear attack pattern can severely deteriorate the HVAC's power state deviation. For this reason, the DR community's total power deviation can also be severely deteriorated. As seen in Fig. 6(b), the total power of DR community can only be regulated to approximately 2473.9 kW, and the response gap becomes even more severe than scenario 1, reaching about 529.6 kW. However, by benefiting from our proposed ASD control, the deviations can be eliminated in a short period of time, and the response speed is approximately 26.7 s. This implies that the proposed control can effectively counteract the adverse impact caused by linear attacks.

Moreover, it can be seen from Fig. 6(c) that the average CR can only reach about 29.18% under linear attack, which implies the cyber-attack can lead to severe dysfunction in the DR community. However, by applying the proposed ASD control, all these undesirable results caused by the attack can be eliminated quickly, and the average CR can be recovered to about 99.67%.

In addition, as can be seen in Fig. 6(d), two HVACs have exceeded the customer's tolerated cold limit (black part), which implies the indoor comfort has been compromised severely. Nevertheless, the thermal state can gradually restored to the comfortable temperature region by using the proposed ASD control.

This scenario illustrates that the proposed ASD control can still protect the DR community from linear attacks.

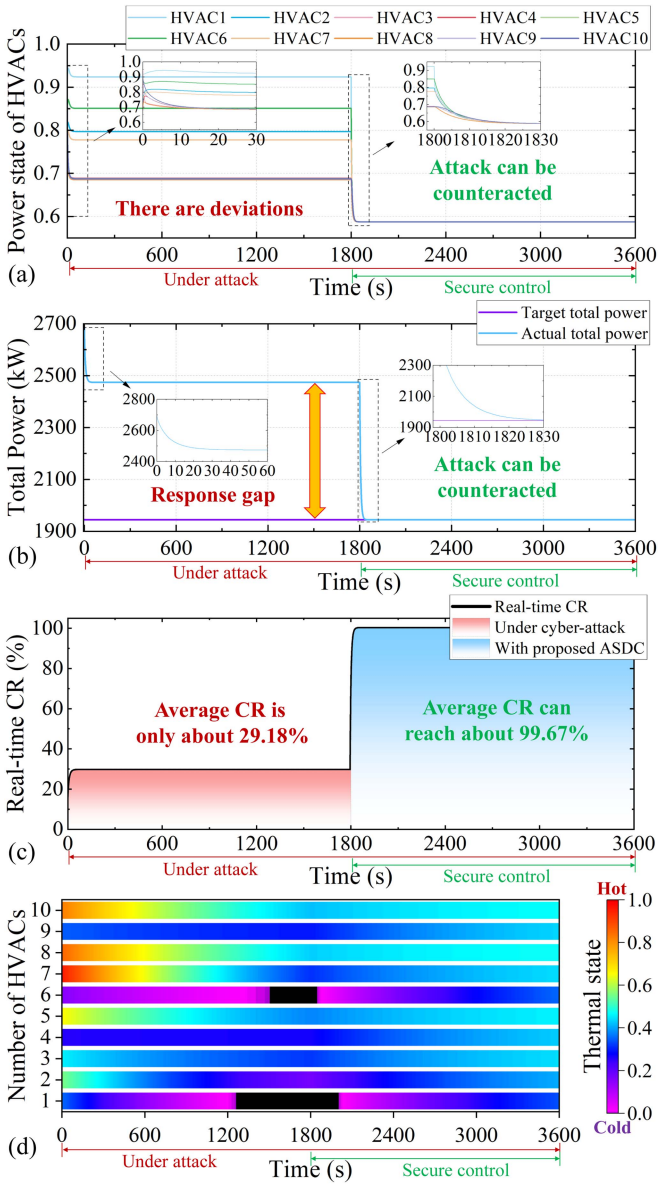

Fig. 6. Performance of the HVAC-based DR under linear attack and efficacy of the proposed ASD control: (a) power state of each HVAC; (b) DR community's total power; (c) real-time CR of DR; and (d) thermal state.

E. Scenario 4: Performance With Non-Linear Attack Pattern

In scenario 4, we consider the non-linear attack pattern described in (13). It is noteworthy that the non-linear attack signal can be arbitrary. Without loss of generality, a cosine function with an amplitude of 0.1 and a frequency of $1/200\pi$ is tested as a non-linear attack signal. The test results are shown in Fig. 7.

As illustrated in Fig. 7(a), under the non-linear attack, the HVACs not only suffer from power state deviations, but never reach consensus convergence and have serious power fluctuations. This is because the injection data of non-linear attack pattern is continuously changing over time, preventing the HVAC's power state from reaching a steady state. As a result, the DR community's total power also suffers from serious power fluctuations over time and can never be stabilized at the target total power required by the power system, as shown in Fig. 7(b). However, by applying the proposed ASD control, the adverse impacts caused by the non-linear attack are eliminated promptly, and the response speed is approximately 18.7 s. Consensus convergence of each HVAC's power state is regained, and the target total power of DR community is reached again, which demonstrates the proposed ASD control can withstand the non-linear attack with arbitrary patterns.

From Fig. 7(c), it can be seen that the real-time CR also fluctuates under the non-linear attack, and the average CR can only reach 89.10%. After entering the secure control stage with the proposed ASD control, the real-time CR is recovered quickly without further fluctuation, and the average CR can reach 99.91%, even under non-linear attacks.

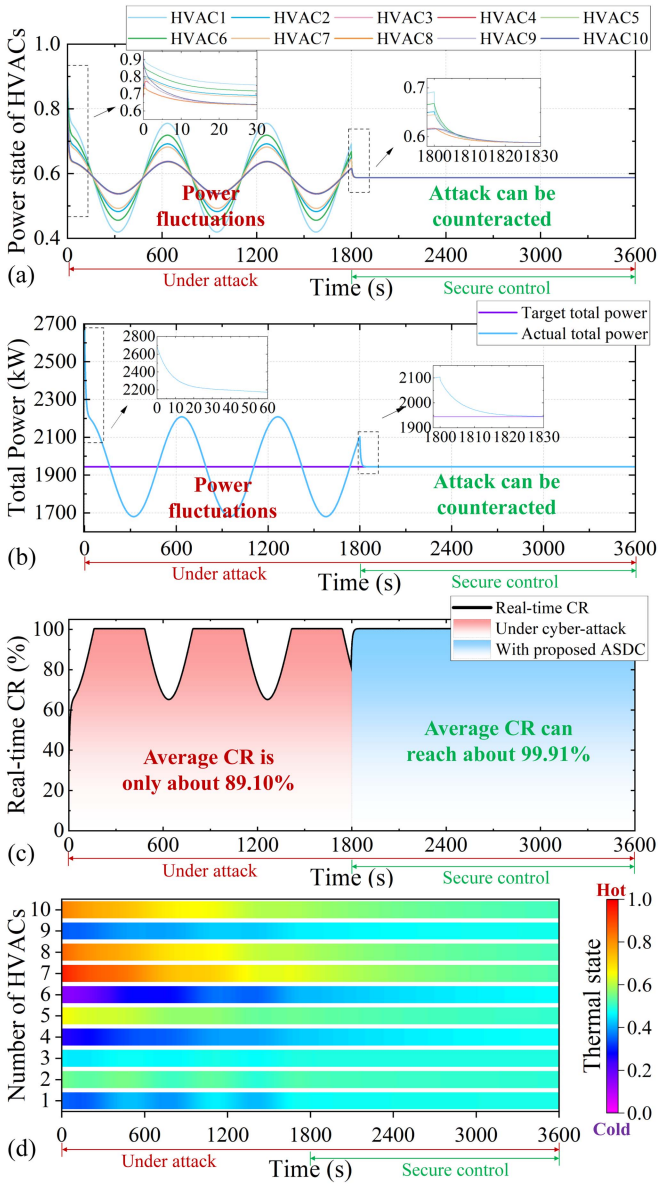

Fig. 7. Performance of the HVAC-based DR under non-linear attack and efficacy of the proposed ASD control: (a) power state of each HVAC; (b) DR community's total power; (c) real-time CR of DR; and (d) thermal state.

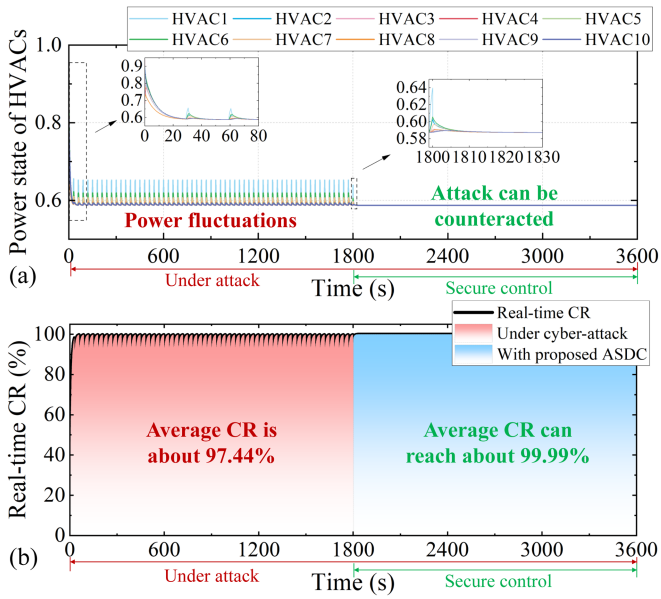

Fig. 8. Performance of the HVAC-based DR under impulsive attack and efficacy of the proposed ASD control: (a) power state of each HVAC and (b) real-time CR of DR.

TABLE III
COMPARATIVE ANALYSIS OF DR IN DIFFERENT SCENARIOS

Scenarios		$\text{Index}_{\text{conv}}$	Average CR
S-1	without ASDC	No	64.25%
	with ASDC	Yes	99.84%
S-2	without ASDC	No	96.39%
	with ASDC	Yes	99.63%
S-3	without ASDC	No	29.18%
	with ASDC	Yes	99.67%
S-4	without ASDC	No	89.10%
	with ASDC	Yes	99.91%
S-5	without ASDC	No	97.44%
	with ASDC	Yes	99.99%

As shown in Fig. 7(d), the thermal state also fluctuates with the dynamic injection data under the non-linear attack. Also, the proposed ASD control can help HVACs avoid such thermal state fluctuation caused by power fluctuation and ultimately stabilize gradually.

F. Scenario 5: Performance With Impulsive Attack Pattern

In scenario 5, we involve a minor amplitude and high-frequency impulsive attack pattern (difficult to detect) to validate the efficacy of the proposed ASD control. The test results are illustrated in Fig. 8.

As shown in Fig 8(a), under impulsive attack, the HVACs in the DR community cannot reach consensus convergence on power state and have serious power fluctuations. However, by applying the proposed ASD control, the adverse impacts caused by the impulsive attack can be eliminated promptly, about 5.65 s. This implies that the proposed ASD control can effectively withstand minor amplitude and high-frequency attacks that are difficult to detect.

Moreover, from Fig 8(b), it can be observed that there are serious fluctuations in the real-time CR under attack, and the average CR is about 97.44%. However, with the proposed ASD control, the undesirable fluctuates can be eliminated quickly, and the average CR can even reach 99.99%. That means that by benefiting from our proposed control, DR tasks can be almost entirely accomplished, despite impulsive attacks.

G. Comparative Analyses for Scenarios With Various DCAs

In this part, a comparative analysis is illustrated in Table III to show the results in different scenarios considering DCAs with various attack patterns, especially including the index of convergence and average CR index. In Table III, the $\text{Index}_{\text{conv}}$

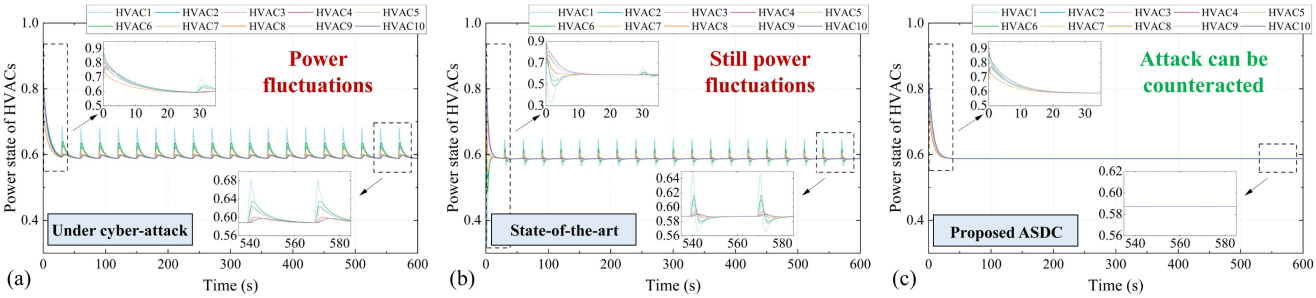

Fig. 9. Performance comparison of power state of each HVAC in the DR community under impulsive attack: (a) under cyber-attack with no action; (b) with the state-of-the-art; (c) with the proposed ASD control.

indicates whether the total power of DR community can converge to the power system's requirement. A 'Yes' indicates that DR can fulfill the operating reserve service required by the power system, while a 'No' indicates the opposite.

As shown in Table III, DR fails to fulfill the regulation requirements of the power system under different attack patterns. However, by utilizing the proposed ASD control, DR can always fulfill the power system's regulation requirements, regardless of the attack patterns. In addition, the average CR is also negatively affected to varying degrees under different attack patterns. Among them, the S-3 scenario is the most severe, with an average CR of only 29.18%. However, by utilizing our proposed ASD control, the average CR can always exceed 99%. That implies that by benefiting from our proposed ASD control, the DCA with arbitrary patterns can be counteracted, and the DR task can always be almost entirely accomplished.

V. Conclusion

DR is vital to power systems since it is conducive to maintaining the power balance between supply and demand. In this work, we formulate the cyber-security problem of DR, specifically addressing DCAs with arbitrary patterns. Moreover, we develop a novel ASD control for DR to defend against arbitrary DCAs comprehensively without relying on detection. Furthermore, the stability and the GUUB convergence are proved rigorously based on Lyapunov theorem. The efficacy of the proposed ASD control is validated in case studies.

The test results demonstrate that different attack patterns are associated with distinct impacts on DR, such as power deviation, delayed response, and power fluctuation, which aligns well with theoretical analyses. However, by using the proposed ASD control, all these adverse impacts can be counteracted, and the DR performance can be improved significantly. For example, under the linear attack, the average CR of DR during the entire test can be increased from 29.18% to 99.67%. Therefore, the proposed ASD control can contribute to the security of DR in harsh cyber environments.

Appendix
Comparison With the State-of-the-Art

In this part, we compare and analyze the defense effectiveness of our proposed ASD control with the state-of-the-art [29] in the impulsive attack pattern scenario, as shown in Fig 9.

From Fig 9(a), it can be seen that under the impulsive pattern DCAs, the HVACs cannot reach consensus convergence on power state and have serious power fluctuations. This implies the DR community's total power also suffers from serious power fluctuations and cannot be stabilized at the target total power required by the power system.

As can be observed from Fig 9(b), the state-of-the-art solution cannot help HVACs eliminate power fluctuations, and HVACs still fail to achieve consensus convergence on power state. This means that the DR community still fails to meet the power system's requirements, even with the state-of-the-art solution.

The performance of our proposed ASD control under impulsive attack is illustrated in Fig 9(c). It can be seen that the consensus convergence of each HVAC's power state can be restored, and the adverse attack impacts can be counteracted.

This comparative analysis illustrates that, compared to the state-of-the-art, the proposed ASD control can effectively defend against DCA with impulsive patterns. This implies that the efficacy of the proposed ASD control outperforms the state-of-the-art, and by benefiting from our method, the DR task can still be accomplished, despite DCAs with arbitrary patterns.

References

[1] N. E. Hultman et al., "Fusing subnational with national climate action is central to decarbonization: The case of the United States," *Nature Commun.*, vol. 11, no. 1, 2020, Art. no. 5255.
[2] H. Li, Z. Ren, A. Trivedi, D. Srinivasan, and P. Liu, "Optimal planning of dual-zero microgrid on an island toward net-zero carbon emission," *IEEE Trans. Smart Grid*, vol. 15, no. 2, pp. 1243–1257, Mar. 2024.
[3] International Renewable Energy Agency, "Renewable capacity statistics 2023" Tech. Rep., 2023. [Online]. Available: https://www.irena.org/Publications/2023/Mar/Renewable-capacity-statistics-2023
[4] P. Wang, Z. Zhang, T. Ma, Q. Huang, and W.-J. Lee, "Parameter calibration of wind farm with error tracing technique and correlated parameter identification," *IEEE Trans. Power Syst.*, vol. 38, no. 6, pp. 5200–5214, Nov. 2023.
[5] H. Hui, Y. Ding, K. Luan, T. Chen, Y. Song, and S. Rahman, "Coupon-based demand response for consumers facing flat-rate retail pricing," *CSEE J. Power Energy Syst.*, early access, Apr. 20, 2023, doi: 10.17775/CSEEJPES.2021.05140.
[6] S. Yang, K.-W. Lao, H. Hui, and Y. Chen, "A robustness-enhanced frequency regulation scheme for power system against multiple cyber and physical emergency events," *Appl. Energy*, vol. 350, 2023, Art. no. 121725.
[7] P. Siano, "Demand Response and smart grids—A survey," *Renewable Sustain. Energy Rev.*, vol. 30, pp. 461–478, 2014.
[8] M. Song, C. Gao, M. Shahidehpour, Z. Li, S. Lu, and G. Lin, "Multi-time-scale modeling and parameter estimation of TCLs for smoothing out wind power generation variability," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 105–118, Jan. 2019.

[9] S. Wang, J. Zhai, H. Hui, Y. Ding, and Y. Song, "Operational reliability of integrated energy systems considering gas flow dynamics and demand-side flexibilities," *IEEE Trans. Ind. Informat.*, vol. 20, no. 2, pp. 1360–1373, Feb. 2024.
[10] Q. Zhou, Z. Tian, M. Shahidehpour, X. Liu, A. Alabdulwahab, and A. Abusorrah, "Optimal consensus-based distributed control strategy for coordinated operation of networked microgrids," *IEEE Trans. Power Syst.*, vol. 35, no. 3, pp. 2452–2462, May 2020.
[11] S. Yang, K.-W. Lao, H. Hui, Y. Chen, and N. Dai, "Real-time harmonic contribution evaluation considering multiple dynamic customers," *CSEE J. Power Energy Syst.*, early access, Apr. 20, 2023, doi: 10.17775/CSEEJPES.2022.06570.
[12] H. Früh, S. Müller, D. Contreras, K. Rudion, A. von Haken, and B. Surmann, "Coordinated vertical provision of flexibility from distribution systems," *IEEE Trans. Power Syst.*, vol. 38, no. 2, pp. 1834–1844, Mar. 2023.
[13] Y. Chen et al., "Distributed self-triggered control for frequency restoration and active power sharing in islanded microgrids," *IEEE Trans. Ind. Informat.*, vol. 19, no. 10, pp. 10635–10646, Oct. 2023.
[14] Y. Chen et al., "Self-triggered coordination of distributed renewable generators for frequency restoration in islanded microgrids: A low communication and computation strategy," *Adv. Appl. Energy*, vol. 10, 2023, Art. no. 100128.
[15] H. Hui, Y. Chen, S. Yang, H. Zhang, and T. Jiang, "Coordination control of distributed generators and load resources for frequency restoration in isolated urban microgrids," *Appl. Energy*, vol. 327, 2022, Art. no. 120116.
[16] X. Zhang, M. Pipattanasomporn, T. Chen, and S. Rahman, "An IoT-based thermal model learning framework for smart buildings," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 518–527, Jan. 2020.
[17] Q. Zhou, M. Shahidehpour, A. Alabdulwahab, and A. Abusorrah, "A cyber-attack resilient distributed control strategy in islanded microgrids," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 3690–3701, Sep. 2020.
[18] "Cyber-attack against ukrainian critical infrastructure," Cybersecurity Infrastructure Security Agency, Tech. Rep. IR-ALERT-H-16-056-01, Jul. 2021. [Online]. Available: https://www.cisa.gov/news-events/ics-alerts/ir-alert-h-16-056-01
[19] E. Bajramovic, D. Gupta, Y. Guo, K. Waedt, and A. Bajramovic, "Security challenges and best practices for IIoT," Gesellschaft für Informatik e.V., Bonn, Sep. 2019, pp. 243–254, doi: 10.18420/inf2019_ws28.
[20] J. Devanny, L. R. F. Goldoni, and B. P. Medeiros, "The 2019 venezuelan blackout and the consequences of cyber uncertainty," *Revista Brasileira de Estudos de Defesa*, vol. 7, no. 2, pp. 37–57, 2020.
[21] J. Hou, F. Teng, W. Yin, Y. Song, and Y. Hou, "Preventive-corrective cyber-defense: Attack-induced region minimization and cybersecurity margin maximization," *IEEE Trans. Power Syst.*, early access, Nov. 15, 2023, doi: 10.1109/TPWRS.2023.3333045.
[22] M. S. Rahman, M. A. Mahmud, A. M. T. Oo, and H. R. Pota, "Multi-agent approach for enhancing security of protection schemes in cyber-physical energy systems," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 436–447, Apr. 2017.
[23] Y. Chen, D. Qi, H. Dong, C. Li, Z. Li, and J. Zhang, "A FDI attack-resilient distributed secondary control strategy for islanded microgrids," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 1929–1938, May 2021.
[24] H. Jahangir, S. Lakshminarayana, C. Maple, and G. Epiphaniou, "A deep-learning-based solution for securing the power grid against load altering threats by IoT-Enabled devices," *IEEE Internet Things J.*, vol. 10, no. 12, pp. 10687–10697, Jun. 2023.
[25] S. Amini, F. Pasqualetti, and H. Mohsenian-Rad, "Dynamic load altering attacks against power system stability: Attack models and protection schemes," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2862–2872, Jul. 2018.
[26] Z. Chu, S. Lakshminarayana, B. Chaudhuri, and F. Teng, "Mitigating load-altering attacks against power grids using cyber-resilient economic dispatch," *IEEE Trans. Smart Grid*, vol. 14, no. 4, pp. 3164–3175, Jul. 2023.
[27] S. Tan, J. M. Guerrero, P. Xie, R. Han, and J. C. Vasquez, "Brief survey on attack detection methods for cyber-physical systems," *IEEE Syst. J.*, vol. 14, no. 4, pp. 5329–5339, Dec. 2020.
[28] J. Hong, H. Hui, H. Zhang, N. Dai, and Y. Song, "Event-triggered consensus control of large-scale inverter air conditioners for demand response," *IEEE Trans. Power Syst.*, vol. 37, no. 6, pp. 4954–4957, Nov. 2022.
[29] S. Yang, K.-W. Lao, Y. Chen, and H. Hui, "Resilient distributed control against false data injection attacks for demand response," *IEEE Trans. Power Syst.*, vol. 39, no. 2, pp. 2837–2853, Mar. 2024.
[30] W. He, W. Xu, X. Ge, Q.-L. Han, W. Du, and F. Qian, "Secure control of multiagent systems against malicious attacks: A brief survey," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 3595–3608, Jun. 2022.
[31] H. Hui et al., "A transactive energy framework for inverter-based HVAC loads in a real-time local electricity market considering distributed energy resources," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 8409–8421, Dec. 2022.
[32] "Cybersecurity for industrial automation & control environments" Schneider Electric,Tech. Rep. 998-2095-04-13-13AR0_EN, 2013. [Online]. Available: https://www.se.com/ww/en/download/document/998-2095-04-13-13AR0_EN/?ssr=true
[33] S. Vitturi, C. Zunino, and T. Sauter, "Industrial communication systems and their future challenges: Next-generation ethernet, IIoT, and 5G," *Proc. IEEE*, vol. 107, no. 6, pp. 944–961, Jun. 2019.
[34] D. Kirovski, F. A. P. Petitcolas, and Z. Landau, "The replacement attack," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 6, pp. 1922–1931, Aug. 2007.
[35] W. He, F. Qian, Q.-L. Han, and G. Chen, "Almost sure stability of nonlinear systems under random and impulsive sequential attacks," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3879–3886, Sep. 2020.
[36] H. K. Khalil, *NonlinearSystems*, 3rd Ed. Englewood Cliffs, NJ, USA: Prentice Hall, 2002.
[37] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[38] Gree Central Air Conditioner Official Website, "Product profile of GREE CVE seires," GREE, Tech. Rep., 2023. [Online]. Available: http://www.glzykt.com/zhanshi/yc.asp
[39] *Performance Rating of Water Chilling Packages Using the Vapor Compression Cycle*, ARI Standard 550/590-2003, Air-Conditioning, Heating, and Refrigeration Institute (AHRI), Arlington, VA, 2003. [Online]. Available: https://www.ahrinet.org/
[40] "Results of meteorological observations," Macao Meteorological and Geophysical Bureau, Tech. Rep., 2023. [Online]. Available: https://www.smg.gov.mo/zh/subpage/348/report/download-pdf
[41] H. Hui, Y. Ding, W. Liu, Y. Lin, and Y. Song, "Operating reserve evaluation of aggregated air conditioners," *Appl. Energy*, vol. 196, pp. 218–228, 2017.
[42] German Transmission System Operators, "Prequalification process for balancing service providers in Germany," Tech. Rep., 2020, Bedingungen FCR aFRR mFRR en. [Online]. Available: https://www.regelleistung.net/ext/download/PQ

Shaohua Yang (Student Member, IEEE) is currently working toward the Ph.D. degree in electrical and computer engineering with the University of Macau, Macao, China. His research interests include cyber-physical security, demand response, and power quality.

Keng-Weng Lao (Senior Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical and electronics engineering from the Faculty of Science and Technology, University of Macau, Macau, China, in 2009, 2011, and 2016, respectively. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, as well as the State Key Laboratory of Internet of Things for Smart City, University of Macau. From June 2017 to June 2019, he was a Research Scholar with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA. His research interests include cyber-physical security, renewable energy integration, energy internet of things, smart energy system protection, and smart grid.

Hongxun Hui (Member, IEEE) received the B.E. and Ph.D. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2015 and 2020, respectively. From 2018 to 2019, he was a Visiting Scholar with the Advanced Research Institute, Virginia Tech, Blacksburg, VA, USA, and the CURENT Center, University of Tennessee, Knoxville, TN, USA. He is currently a Research Assistant Professor with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macao, China. His research interests include optimization and control of power system, demand response, and Internet of Things technologies for smart energy.

Yulin Chen (Member, IEEE) received the Ph.D. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2021. From September 2021 to September 2022, he was a Postdoctoral Fellow with the University of Macau, Macao, China. He is currently an Associate Research Fellow with the Hainan Institute of Zhejiang University, Sanya, China. His research interests include distributed control of renewable energy and cyber-physical security with application in smart grid.