Load Altering Attack-Tolerant Defense Strategy For Load Frequency Control System[★]

Chunyu Chen[a], Mingjian Cui[b,*], Xin Fang[c], Bixing Ren[d] and Yang Chen[e]

[a] *School of Electrical and Power Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China*
[b] *Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, Tennessee 37996, USA*
[c] *National Renewable Energy Laboratory (NREL), Golden, Colorado 80401, USA*
[d] *State Grid Jiangsu Electric Power Co., Ltd. Research Institute, Nanjing, Jiangsu 211103, China*
[e] *School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore*

ARTICLE INFO

Keywords:
load frequency control
load altering attack
model-free control
defense strategy

ABSTRACT

Cyber attacks become emerging threats to every information-oriented energy management system. By violating the cyber systems, the hacker can disrupt the security and stability due to the strong coupling between the cyber and physical facilities. In this paper, one type of cyber attacks designated as the load altering attack is studied for the power system frequency control, and corresponding defense strategies are proposed to improve the frequency control performance. Considering the difficulty of the application of model-based controller into large-scale power systems, a novel model-free defense framework is for the first time presented. Under this framework, both active defense and passive defense strategies are designed. The former assumes that the defender has the initiative to learn different attack scenarios. Adaptive defense strategies are implemented using the online attack identification information and off-line trained strategy pool. The latter assumes that the defender passively tolerates various attack scenarios via the pre-trained off-line strategy. Both approaches prove to be effective through validation based on the IEEE benchmark systems. The proposed defense framework and defense strategies can be extended to other energy control systems to enhance their attack tolerance capability.

1. Introduction

The modern power systems are evolving into energy-cyber-physical systems, thanks to the ever-growing communication networks, advanced computation, and intelligence techniques [1]. Under this circumstance, the cyber security of power systems has become an intensively discussed subject due to global security threats caused by the rampant terrorist attacks. As one essential component in the energy sector, the secured control and operation of electrical power systems, in particular, the fundamental power system frequency stability and control [2], is the key to the proper functioning of energy-dependent activities. Therefore, defense strategies must be established to adapt to this new cyber threat environment.

Defense against cyber attacks is an emerging topic for both the transmission and distribution systems [3]. The focal point of existing studies lies in the detection [4] and isolation [5] of the attack signal; but few genuinely consider a complete defense mechanism design, which is indispensable to the robustness of the control system under attack. In the domain of system operation or planning, various attack objectives including the bad data detection (BDD) performance [6], economy [7], load fulfillment [8] are considered in studying different attacks and corresponding mitigation strategies. To elude BDD more economically and stealthily, research concerning the minimal number of attacked meters or cooperative meter attack [9] emerges. Correspondingly, mitigation strategies including optimal sensor placement [10] and measurement variation dynamics analysis [11] are studied. Cyber attacks on the economic operation are usually modeled by bi-level programming: the upper level is associated with the hacker while the lower level is for the economic operation. Under this frame, the attack goal is to construct an effective attack vector to increase the operational cost [12]. Recently, decentralized programming is used to detect cyber attacks based on the convergence criteria [13]. By isolating the unit when the confidence level of neighbor-observation is beyond the threshold, the

[★] This work is supported by Chengdu Guojia Electrical Engineering Co., Ltd (No. NEEC-2019-A03).

[*] Corresponding author

✉ mingjian.cui@ieee.org (M. Cui)
ORCID(s):

convergence is still guaranteed under attack. Besides the economic damage, some hackers aim at unnecessary load curtailment by compromising power flow measurements known as load redistribution (LR) attack [14]. To quantify the influence of an LR attack on the long-term supply reliability, a holistic evaluation model is proposed in [15]. A dynamic sequential-mining algorithm is used to extract vulnerable branch sequences under LR attack [16].

Since modern power grids are evolving into integrated cyber-physical systems, coordinated cyber-physical attacks are thus investigated to analyze more sophisticated joint attack mechanisms. False data injection attack (FDIA) is coordinated with physical attacks to mask the undetectable transmission line outages [17]. A coordinated false data injection and load altering attack (LAA) is studied for various attack objectives [18]. A two-layer defense model is proposed to optimize the deployment of defense resources using the zero-sum game algorithm [19]. A tri-level optimization model is built to formulate the coordinated attack scenario where the hacker attacks the firewall at the control centers and field devices simultaneously [20].

In this paper, the focal point is the defense strategy against LAA on the secondary frequency control or load frequency control (LFC), which is the fundamental element of the energy management system [21]. By changing the volume of unsecured loads, the operation and control systems are affected by LAAs. In respect to LFC, an LAA could be achieved by manipulating unsecured load integrators (LIs). Unlike dispersed and small-scale residential loads, LIs are integrated groups of residential consumers or large industrial consumers who are expected to participate in demand-side response program. An LAA against unsecured LIs could significantly affect the frequency stability by considering the volume of the load alteration [22].

There exist abundant research studies about cyber-physical attacks on frequency control systems, ranging from attack strategy analysis [23] to defense strategy design [24]. The cyber attack impact of the automatic generation control system is studied in [25]. Besides the ‘dumb’ attack behaviors, some researchers study how to design ‘intelligent’ schemes to optimize specific attack objectives [26]. Cyber attack detection in the frequency control system aims to distinguish the normal frequency excursions with the compromised ones [27]. Recently, game techniques are used to model the mutual interactions between the attacker and defender [28]. Cyber attack-tolerant power system frequency controllers, regarded as the fundamental of the defense strategy in power system frequency control, have been investigated via model-based ones before. A novel simultaneous input and state estimation algorithm is used to detect and compensate for FDIA attacks [29]. An unknown input observer is used to identify cyber attacks against LFC [30]. However, drawbacks of model-based controllers, such as requirements for special conditions and availability of complete model information, may hinder their usage in the real-life application. The cyber attack-tolerant controller is analogous to the resilient operation strategy [31]; both aim to achieve satisfactory control or operational performance under unexpected abnormal disturbances. The difference lies in that the former belongs to the real-time control and the latter is some time-ahead optimal energy management. It determines that resilient optimization methods in the operation domain cannot be used for the control problem considered herein.

Distributed algorithms are widely researched for power system operation due to its flexibility [32]. And a distributed control system (DCS) is a promising technique for secure power system control [33]. With the aid of the consensus protocol, the cyber attack can be detected if the consensus criteria are not satisfied. Specifically, the distributed security observer is designed to analyze, detect and even mitigate attacks. Nevertheless, influences from different control loops might deteriorate the overall performance, and the coordination mechanism should be carefully devised. Also, the extra cost of establishing the DCS center cannot be ignored. Therefore, a reliable and secure controller that can tolerate or defend an LAA in the centralized control mode might be preferred for the transmission systems.

Inspired by fault-tolerant control (FTC) in the control theory branch [34], novel LFC schemes under LAA are investigated by borrowing the idea of FTC and reforming it to better suit the needs herein. Model-free methods using reinforcement learning (RL) are exploited to emancipate the controller from the complex model. RL has been widely researched in energy management of various energy systems. A double Q-learning-based management strategy is designed for the hybrid fuel cell and battery propulsion system to minimize the operational cost [35]. An RL-based energy management strategy is designed for the hybrid construction machinery [36]. With the aid of deep neural networks, deep Q-network (DQN) is further developed for RL-based energy management strategies, which can tackle large scale learning tasks considering various operating conditions[37]. In this paper, both active defense (AD) and passive defense (PD) strategies are designed via RL for complex detailed power system models, which are different from the reduced-order simple plant model in theoretical control studies. As with an active FTC, an AD strategy adapts to specific learned LAAs online by ‘actively’ adjusting the off-line established reconfiguration mechanism; while different LAAs are pre-considered at the design stage of a PD strategy, and it ‘passively’ tolerates the unknown LAAs in the online execution phase.

The proposed strategies are completely decoupled from the power system model, which only serves as an 'environment' where the external observation is used to learn the 'optimal' policy. Therefore, the practicality and performance are significantly improved. Specifically, the main contributions include:

- A novel defense framework is for the first time proposed for the LFC system under LAA to attenuate its influence, and the proposed defense framework can also be extended to the defense of other energy control systems.
- A novel model-free AD strategy is designed under the circumstance where the defender has the initiative to learn different LAAs and uses the learned information for the attack attenuation. More specifically, a composite LAA signal estimation and feedforward compensation-based defense strategy is presented.
- A novel model-free PD strategy is designed under the circumstance where the defender passively tolerates different LAAs by enhancing the system redundancies. More specifically, a PD strategy using the deep Q network (DQN) technique is presented.

The remainder of the paper is as follows: Section 2 gives the basic backgrounds of model-free defense strategies. Section 3 presents a detailed design procedure of a model-free AD-based LFC scheme. Section 4 discusses a model-free PD-based LFC scheme. Numerical analyses are demonstrated in Section 5 and 6 to verify the validity. Eventually, conclusions and future work are addressed in Section 7.

2. Preliminaries

First, we briefly address basics of a model-free LFC defense strategy, including its working principle and superiority (contribution). It lays the foundation for design procedures in Sections 3 and 4.

2.1. Load Altering Attack on Load Frequency Control

With the integration of information and communications technology (ICT), the demand side management (DSM) is becoming an undeniably dominant force of flexible energy supplies. Though the capacity of power supplies is enhanced by DSM programs, ICT-induced vulnerabilities from cyber spaces could disrupt the normal operation of DSM programs and then the system security. Each participator depends on external signals, which are dispatched from the DSM center, to adjust the energy demand. An LAA can be treated as an ill-intentioned energy demand alteration [38]. Fig. 1 demonstrates the complete LAA process. The attacker first exploits the vulnerabilities of communication systems to infiltrate the DSM center, then external signals sent to participators are controlled. The attacker might also control external signals directly through invading communication channels between the DSM center and participators. With fully-controlled external signals, the attacker can arbitrarily alter the consumption of large industrial participators, leading to the abrupt power imbalance and system instability. From the perspective of power flow analyses, the mathematical representation of an LAA is:

$$P_{is} + d = U_i \sum\nolimits_{j\in\mathcal{N}_i} U_j \left(G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}\right), \forall i \in \mathcal{N}_e \tag{1}$$

where i is the index of the unsecured load bus which connects with participators; $\mathcal{N}_e$ is the set of unsecured load; $j \in \mathcal{N}_i$ represents the neighboring bus of load bus i; U is the magnitude of the voltage; θ_{ij} represents the phase angle difference between load bus i and j; G_{ij} and B_{ij} represent the real and imaginary part of the admittance between load bus i and j. d represents the load alteration in an LAA.

In the context of LFC, by combining the power equations of buses that all assume the form of (1) and the dynamic model of units, the mathematical formulation of an LAA on LFC can be expressed as:

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{f}\left(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{d}\right) \\ y = \boldsymbol{g}\left(\boldsymbol{x}\right) \end{cases} \tag{2}$$

where $\boldsymbol{x}$, y, and $\boldsymbol{u}$ represent the vector of power system states (unit angle, the voltage behind the transient reactance, etc.), the output (system frequency), the control input (governor reference value of the LFC-participating units), respectively. $\boldsymbol{f}$ and $\boldsymbol{g}$ are algebraic equations; $\boldsymbol{d}$ represents the vector of LAA signals.

Based on (2), the goal of LAA-tolerant controller design is to calculate u, such that the impact of LAAs on the system frequency y can be alleviated, i.e., the system frequency can remain the small neighborhood of the equilibrium.

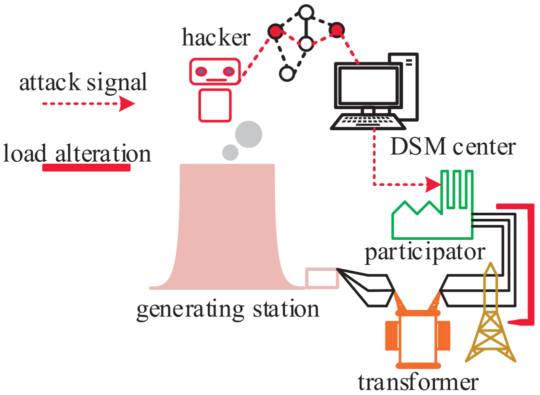


Figure 1: Schematic diagram of load altering attack

2.2. Model-Free Load Frequency Control

By the algebraic derivation operator, the simple linear ultra-local model can replace the complex nonlinear system model (2) [39]:

$$\underbrace{\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{d}) \\ y = \boldsymbol{g}(\boldsymbol{x}) \end{cases}}_{nonlinear\ \ model} \rightarrow \underbrace{y^{(\upsilon)} = \hat{\boldsymbol{\alpha}}\boldsymbol{u} + \hat{\boldsymbol{F}}}_{ultra-local\ \ model} \tag{3}$$

where $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{u}$, $\boldsymbol{d}$, $\boldsymbol{g}$ and $\boldsymbol{f}$ have the same meaning as (2); υ is the derivation order. The 'ultra-local' means that the model is simplified as a weighted sum of an aggregated nonlinear disturbance term and the control input, independent from the original 'global' model information; $\hat{\boldsymbol{F}}$ and $\hat{\boldsymbol{\alpha}}$ are the estimation of lumped nonlinear terms and the control input coefficient. Based on the ultra-local model, strategies which are suitable for low-complexity small-scale system, including intelligent PID, fuzzy system control and sliding mode control, can be used to achieve model-free control (MFC).

The MFC herein completely disintegrates with the model (either the original or the ultra-local one). The controller rather learns from the system responses than the model dynamics for control command generation. The transformed ultra-local model in (3) is not required, and the MFC is based upon the original untouched system, thus solving the problem of obtaining υ in (3) from complex high-order power systems.

2.3. Operating Principle of Active or Passive Defense-Based Load Frequency Control

In this section, the principle of AD and PD-based LFC, which guides Section 3 and 4, is presented for a better understanding of the working principle.

2.3.1. Operating Principle of Active Defense-Based Load Frequency Control

Since an LAA can be equivalent to a fault in FTC, AD-based LFC is analogous to active fault-tolerant control (AFTC), the aim of which is to achieve reliable performance through fault compensation. Technically, AFTC uses the information of fault detection and diagnosis (FDD) to supervise the reconfiguration mechanism. In the context of AD-based LFC under LAA, it means that the controller should collect the diagnosis information of LAAs (d in (2)) before reconfiguring the original LFC controller. Inspired by the architecture of AFTC in [40], the schematic diagram of AD-based LFC is shown in Fig. 2. As can be seen, this AD strategy actively exploits the system dynamics to design the attack detection and diagnosis (ADD) module. It can thus obtain the quantitative information of d through dynamic observer, parameter estimation, or parity techniques [41]; hence, it pertains to the active control. However, model-based strategies are barely achievable for complex power system systems. Instead, model-free methods are used

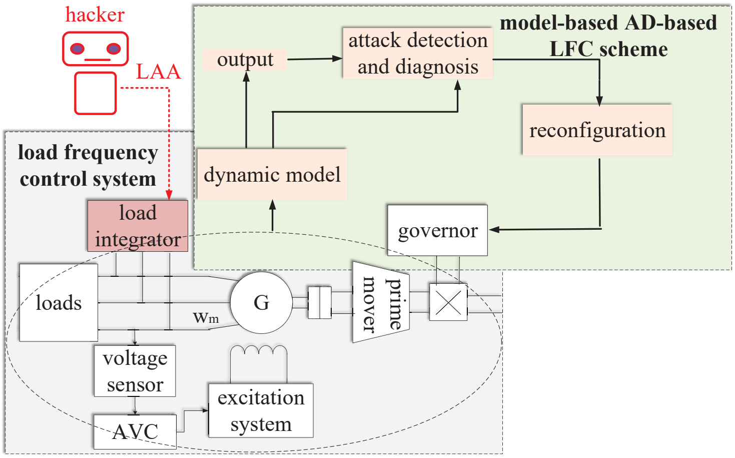


Figure 2: Schematic diagram of AD-based LFC under LAA

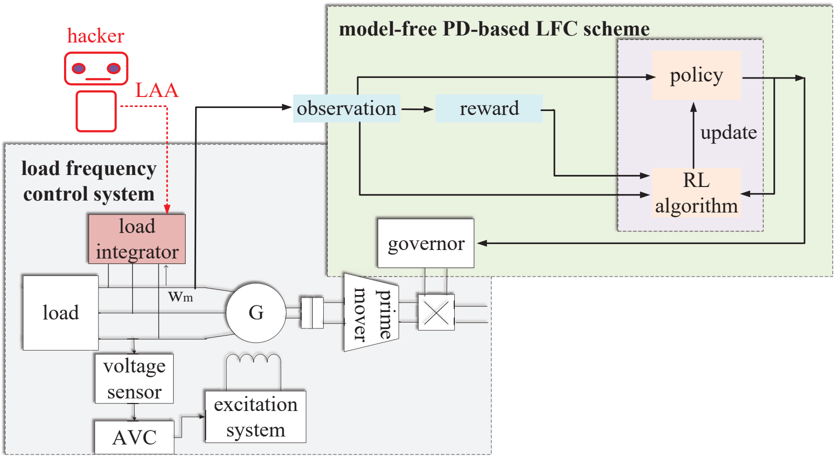


Figure 3: Schematic diagram of PD-based LFC under LAA

to achieve ADD herein; in this sense, the AD-based LFC can be regarded as a special type of model-free active FTC scheme.

2.3.2. Operating Principle of Passive Defense-Based Load Frequency Control

Though the control goals are the same, unlike model-free AD-based LFC, PD-based LFC does not actively perform ADD, and the controller reconfiguration mechanism is not required. Instead, it tolerates LAAs by using system redundancies as passive FTC does. The controller teaches itself to search for the 'best' strategies to 'minimize' the LAA impact, in the manner of pure data-driven exploration-and-exploitation tactics such as reinforcement learning (RL). And the redundancies herein is interpreted as the generalization ability of the neural network-based policy agents, which can produce good outputs for LAAs not encountered in the training process. Fig. 3 demonstrates the philosophy of PD-based LFC.

Based on Section 2.1, the benefits of model-free defense strategies are as follows:

- Compared with the model-based controller, it does not require the dynamic model information to achieve attack diagnosis, which would dramatically increase the cost and complexity of the controller. Therefore, it is appealing to large-scale power system control.

- Compared with MFC using ultra-local models, it does not require additional parameter information as the reduced-order ultra-local model does. For example, precise υ, $\hat{F}$, and $\hat{\alpha}$ in (3) are demanded in MFC based on ultra-local models.
- The proposed MFC-based defense strategy uses the RL technique to obtain the 'optimal' strategy. Once trained, every control action is considered 'optimal' concerning the maximization of the long-term expected return (which quantifies the control performance). It automatically executes according to the current system states, thus achieving the quick control velocity and stability simultaneously by avoiding the latency in closed-loop feedback control.

3. Model-Free Active Defense-Based Load Frequency Control Through Attack Detection and Diagnosis

Based on the rules in Section 2.3.1, a novel composite model-free AD strategy for LFC is designed herein.

3.1. Basic Framework of Active Defense-Based Load Frequency Control

As can be seen from (2), by obtaining the estimation $\hat{d}$ and taking control actions $u = f(\hat{d})$ promptly, LAA impacts could be significantly attenuated. As with model-based AD, $\hat{d}$ is achieved through ADD. Nevertheless, the ADD herein is rather a data-driven than model-based one (e.g., attack observer using system dynamics) because complex nonlinear models considered herein are unsuitable for model-based ADD.

The data-driven ADD means that it demands a specific period to extract time-domain information of d by collecting the frequency transients. The execution time of ADD plus the response time of pure feedback control would extend the total recovery period. Therefore, RL-based MFC is adopted after ADD, in the manner of feedforward compensation. The procedural form is shown in Algorithm 1.

Algorithm 1 Composite AD-based LFC using RL and ADD

```
Initialization: Set current time $t$, lag operator index $k_1$ and lookahead index $k_2$, obtain $\mathcal{W} = \begin{bmatrix} \omega(t) & \omega(t-1) & \cdots & \omega(t-k_1) \end{bmatrix}$. Set the lower and upper bound of FD $d \in (d_l, d_u)$.
Step 1. Establish the discrete set $\mathcal{D} = \{d\}$ by using the sampling method, with each element $d_i$ representing a specific LAA scenario.
Step 2. For each $d_i$, train the RL agent to obtain the corresponding $\pi_i^*$
Step 3. Predict false power injection $d$ at the next time point $t + k_2$ using the regression model $\hat{d}_{t+k_2} = Regress(\mathcal{D}, \mathcal{W})$.
Step 4. Search for the nearest (to $d(t + k_2)$) element $d_o$ from $\mathcal{D}$, obtain the trained policy $\pi_o^*$.
Step 5. Calculate the control error $y_e$ and feedback law $u_f = k(y_e)$.
Step 6. Calculate the control input as $u = u_f + \pi_o^*$.
```

Algorithm 1 is generally categorized into three parts:

- **Part I: RL-based AD** This is the core of the optimal AD strategy obtained from Step 1 to 2 for different attack scenarios.
- **Part II: d prediction & optimal strategy** The predictive information of LAAs (in Step 3) offered can be used to find the suitable pre-trained strategy (from Part I) through scenario matching (in Step 4) in advance, thus synthesizing the real-time control.
- **Part III: feedback control-based compensation** The feedback control (in Step 5) compensates for the errors due to the matching and prediction uncertainties (in Part II).

3.2. Part I: Reinforcement Learning-Based Active Defense

From Step 1 and 2 in Algorithm 1, it is learned that RL-based AD predetermines optimal AD strategies for a variety of LAAs characterized by different d. Then the suitable candidate among the strategy pool can be used after scenario matching.

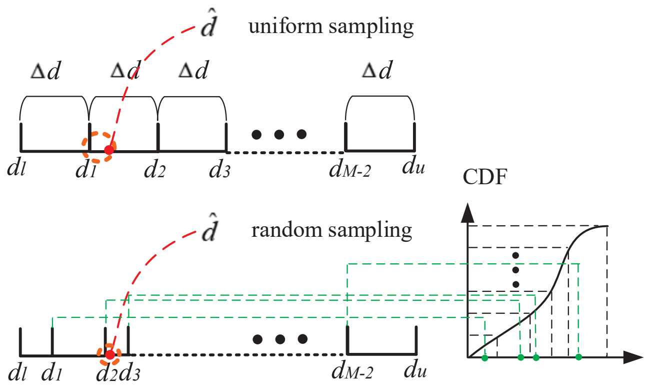


Figure 4: Sample generation based on the probability distribution of the LAA signal

3.2.1. *Sample Generation*

Step 1 in Algorithm 1 aims to generate different LAA signal samples for training. Two sampling methods can be used for sample generation: 1) uniform sampling and 2) random sampling. The latter (e.g., Latin hyperbole sampling) is distribution information-dependent. Supposing that M samples are required, these two methods can be generalized as the schematic in Fig. 4. Random sampling can reduce unnecessary samples that fall in the low probability density areas (e.g., the tail) supposing that the distribution of d is known, giving more samples which the attacker would choose with high likelihood. Hence the 'minimal distance' between the predicted $\hat{d}$ is much smaller in random sampling than that in uniform sampling. For example, the distance between d_2 and $\hat{d}$ is much smaller than the distance between d_1 and $\hat{d}$ in Fig. 4, which is desired to enhance the robustness of the AD strategy.

3.2.2. *Off-Policy Active Defense Strategy*

Off-policy RL is adopted due to the following reasons:

- On-policy RL is more conservative in that it follows the same policy derived from state-action values, and it ignores the other possible 'better' policies, which is avoided in off-policy RL by the maximization operator.
- The core of RL-based AD is to use off-line trained strategies for online execution, which is following the off-policy mode.

Instead of using primitive Q learning [42], actor-critic (AC) is used to handle the continuous space problem. For a specific d_i from the set $\mathcal{D} = \{d\}$ obtained by sample generation, the procedural form of AC-based AD strategy is shown in Algorithm 2.

Algorithm 2 AC-based AD strategy

```
Initialization: Set Critic $Q(a, s; w)$ and Actor $\pi(a|s; \theta)$. Set the sampling time $T_s$. Set the learning coefficient $\alpha$ and $\beta$. Set the COI frequency $\omega_{COI}$ the state $s$ of the simulation environment, which is characterized by a differential-algebraic equation-based dynamic model [24]. Set reference power of the governor the action $a$ of the simulation environment.
Step 1. While Not Converged do
    Step 2. Observe current state $s_t$ and obtain the action based on the Actor $a_t \sim \pi\left(.|s_t; \theta_t\right)$; perform this $a_t$ on the governor and observe the new $s_{t+1}$ after $T_s$, and compute the reward:
        $r_t = -100\left|\Delta\omega_{COI}\right| - 25\left|\frac{\partial\omega_{COI}}{\partial t}\right|,$
    Step 3. Obtain $a_{t+1} \sim \pi\left(.|s_{t+1}; \theta_t\right)$. Evaluate the Critic $Q_t = Q\left(s_t, a_t; w_t\right)$ and $Q_{t+1} = Q\left(s_{t+1}, a_{t+1}; w_t\right)$. Compute the TD error:
        $\delta_t = Q_t - \left(r_t + \gamma Q_{t+1}\right)$
    Step 4. Compute the gradient and update the Critic:
        $w_{t+1} = w_t - \alpha\delta_t \frac{\partial Q\left(s_t, a_t; w\right)}{\partial w}|_{w=w_t}$
    Step 5. Compute the gradient and update the Actor:
        $\theta_{t+1} = \theta_t + \beta\delta_t \frac{\partial \log\left(a_t|s_t, \theta\right)}{\partial\theta}|_{\theta=\theta_t}$
end while
Step 6. Execute $a_t \sim \pi\left(.|s_t; \theta_t\right)$ with the well-trained Actor.
```

3.3. Part II: d Estimation & Optimal Strategy

To obtain $\hat{d}$ in Fig. 4, the estimation or prediction via regression must be performed. Sufficient reaction time should be preserved to match $\hat{d}$ with the samples in Fig. 4 before executing the suitable off-line policies.

Since LFC essentially solves the small-signal stability problem, for the convenience of analysis, the system under LAA is expressed by the small-signal model (Z-domain transfer function):

$$D(Z) = \frac{a_m z^m + a_{m-1} z^{m-1} + \cdots + a_0}{b_n z^n + a_{m-1} z^{n-1} + \cdots + b_0} W(Z) \tag{4}$$

where $D(Z)$ represents d in Z-domain, $W(Z)$ represents the COI frequency of the system in Z-domain. (4) can be rewritten as:

$$\begin{array}{l} D(Z) = \frac{g_{m-n} z^{m-n} + g_{m-n-1} z^{m-1-n} + \cdots + g_n z^{-n}}{1 + c_1 z^{-1} + \cdots + c_n z^{-n}} W(Z) \\ D(Z) + \cdots + c_n D(Z) z^{-n} = g_{m-n} W(Z) z^{m-n} + \cdots + g_n W(Z) z^{-n} \\ d(k) + \cdots + c_n d(k-n) = g_{m-n}\omega(k+m-n) + + \cdots + g_n \omega(k-n) \\ d(k) = f_1(d(k-1), \cdots, d(k-n)) + f_2(\omega(k+m-n), \cdots, \omega(k-n)) \end{array} \tag{5}$$

As can be seen, $d(k)$ is a weighted sum of d and ω at different discrete instants. It means that d and ω in the past several discrete instants can be used to predict the future d by estimating f_1 and f_2 (which can be nonlinear functions in the context of the large-signal model). This prediction problem belongs to the regression, which can be solved by classic regression and advanced deep learning methods [43].

The brief procedural form of the regression is summarized in Algorithm 3.

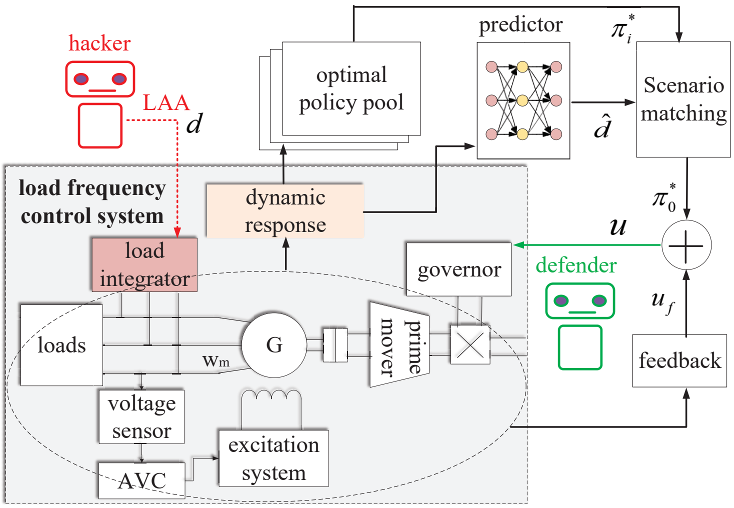


Figure 5: Load altering attacks on Kundur's 4-unit-13-bus system

Algorithm 3 regression-based d estimation

```
*Initialization*: Set total simulation time $T$; set the sampling index $k$; set the training/test ratio $N_1/N_2$; set the time step $T_s$; set the sampling $T_a$; set the maximal number of instances $n_{max}$.
**Step 1. for** $t \leq T$ **do**:
    Simulate the model with step size $T_s$ by randomly choosing $d_0$ between the predetermined $(d_{min}, d_{max})$; store $\omega_0$ and $d_0$ at every discrete time instant.
**end for**
**Step 2. for** $i = 1 : length(t)$
    if $mod(i-1, ceil(T_a/T_s)) == 0$
        $\omega(k) = \omega_0(i)$; $d(k) = d_0(i)$,
**end for**
**Step 3. While** $j < n_{max}$ **do**:
    set the $j^{th}$ instance $\phi_i = [x_i, y_i]$ with $x_i$ the input and $y_i$ the output.
**end for**
**Step 4.** Disorder $\{\phi_i\}$ randomly; obtain the training data $\Phi_{tr}$ and test data $\Phi_{te}$ based on the training/test ratio $N_1/N_2$.
**Step 5.** Train the regression-based predictor $P(x,\theta)$ by using $\Phi_{tr}$. Test $P(x,\theta)$ by using $\Phi_{te}$.
**Step 6.** Simulate a new LAA signal $d$ and estimate it using the trained $P(x,\theta)$.
```

Step 2 in Algorithm 3 is used to reduce the volume of the training data. With the aid of Algorithm 3, $\hat{d}$ at the future $t+k_2$ can be obtained. The optimal policy π_o^* for the nearest sample $d_o = \min_{d_i} \left|\hat{d} - d_i\right|$ is thus executed at $t+k_2$.

3.4. Part III: Feedback Control-Based Compensation

There could exist errors between the real d and predicted $\hat{d}$. Therefore, corrective action (feedback control) could be performed to counteract the estimation errors.

The control law $u_1 = k(\hat{d})$ obtained in Section 3.3 is dependent upon $\hat{d}$, which might be slightly different from the real d. Without increasing the complexity and cost, a feedback correction term $u_2 = k_f(\omega - \omega_0)$ is used for error reduction. The diagram of the composite controller considering feedback control is shown in Fig. 5. It should be mentioned that this feedback correction module can also be removed for the simplicity of implementation. The pure feedforward compensation-based AC agent can still guarantee the frequency quality.

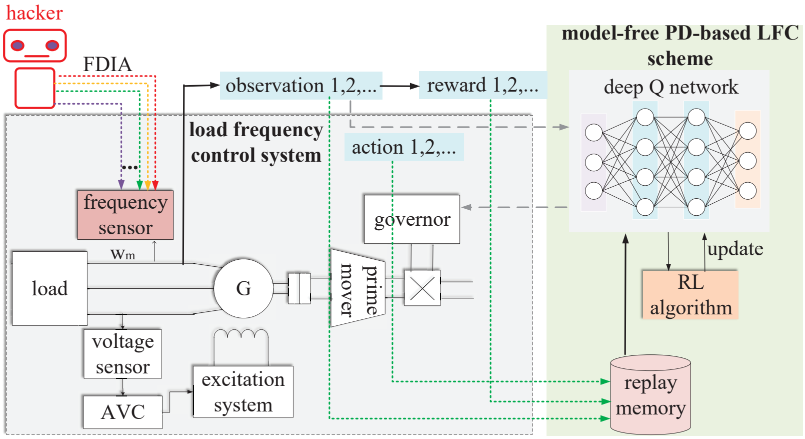


Figure 6: PD-based load frequency control using DRL

4. Model-Free Passive Defense-Based Load Frequency Control Through Deep Reinforcement Learning

The AD strategy in Section 3 relies on d in the online operation. Based on Section 2.3.2, its PD counterpart that requires no knowledge of d is designed herein.

4.1. Basics of Passive Defense-Based Load Frequency Control

Inspired by passive FTC [40], a PD-based LFC scheme is proposed. All potential attack scenarios are considered at the design stage. The ADD in AD strategies is not required, relieving the burden of repeated controller reconfiguration. Compared with AD strategies in Section 3, it appeals to scenarios where the minimal human intervention is expected during the operation.

The PD-based LFC still uses RL; nevertheless, unlike AC-based Algorithm 3, PD strategies consider various LAAs simultaneously, possessing generalization ability for unknown scenarios by using deep reinforcement learning (DRL). The schematic diagram in Fig. 6 shows that it uses replay memory to store data batches deriving from multiple transitions during different episodes under different LAAs (indicated by dashed lines with different colors in Fig. 6). Then batch learning is performed using certain loss functions to update DQN hyperparameters, which produce the 'optimal' strategy for an unknown LAA scenario.

4.2. Deep Reinforcement Learning-based Optimal Control

The observation state used for reward computation is COI frequency deviation ω_{COI}. The procedural form of Fig. 6 is summarized in Algorithm 4. *Con* in Step 3 represents the stop criterion. If the COI frequency exceeds specific boundary values, it would be heavily punished in the reward with a large negative coefficient of -4000. The benefit of this setting is to avoid dumb exploration around the exorbitant action values. The $Mean$ operator is used to smooth the action and the system dynamic responses. As can be seen, Algorithm 4 abandons the predictor design (Algorithm 3) and scenario matching in Algorithm 1, and it is purely off-line trained without any manual intervention. And it can still handle different LAA scenarios due to the generalization ability using DQN. It can be regarded as the advanced version of Algorithm 2, which does not adopt DQN and thus cannot handle multi-scenario by a single AC network.

From Sections 3.1 and 4.2, it can be seen that neither the AC-based AD strategy nor the DQN-based PD strategy violates the principles concerning the real timeliness of LFC. In the AC-based AD strategy, each pair of actor and critic networks is tuned offline for a specific attack scenario characterized by an LAA signal. The regression network, which maps the COI frequency and the LAA signal, is also offline trained. During the online operation, for an unknown LAA, its value is first estimated by the regression network. Then its nearest neighbor is obtained from the offline-trained AC networks and used to generate the control command. The time of these processes can be ignored when compared with the control cycle, having no influence on the real timeliness of LFC. Similarly, in the DQN-based PD strategy, hyperparameters of DQN are tuned offline to generate 'optimal' output for various attack scenarios. During the online

Algorithm 4 Deep Q network-based passive defense strategy

```
Initialization: Initialize replay memory $D$; initialize $Q$ with random weights $\theta$; initialize $\hat{Q}$ with random weights $\hat{\theta}$.
Set the maximal episode number $M$, set the maximal sampling time $T$ of each episode.
Step 1. For episode= $1, M$ do
    Step 2. Initialize sequence $\phi_1$ For $t = 1, T$ do
        Step 3. Choose $a_t$ with $\varepsilon-$ greedy algorithm. Execute $a_t$ on the governor, obtain the next sequence $s_{t+1}$ and compute the reward $r_t$:
            $r_t = -40\sqrt{|\omega_{COI}|} - 20|\omega_{COI}|^2 - 4000Con$
        Step 4. Store the transition $(\phi_t, a_t, r_t, \phi_t)$ in $D$, sample random minibatch of transitions from $D$, and set:
            $y_j = r_j + \gamma \max_{a'} \hat{Q}\left(\phi_{j+1}, a'; \hat{\theta}\right)$
        Step 5. Obtain the gradient of $\left(y_j - Q\left(\phi_j, a; \theta\right)\right)^2$ and update $\hat{Q}$ .
    end for
end for
Step 6. For a specific online LAA scenario Execute $a_t = Mean\left(argmax_a \hat{Q}(s, a; \theta)\right)$
```

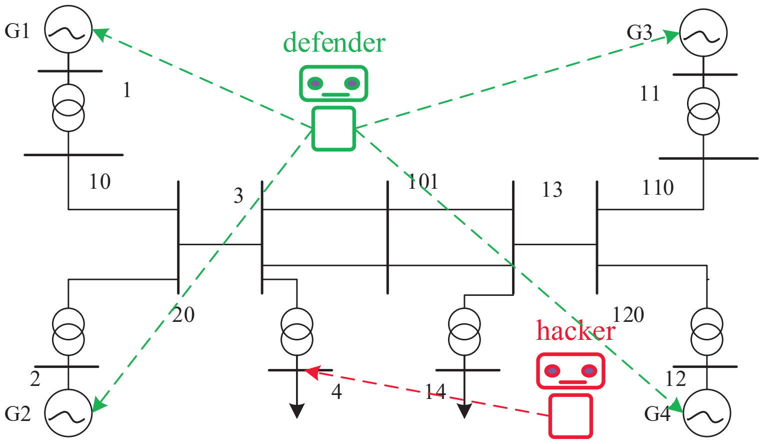


Figure 7: Load altering attacks on the Kundur's 4-unit-13-bus system

operation, DQN uses the frequency information to take less time in computing the compensation. Thus, it does not affect the real timeliness of LFC.

5. Case Studies for the Kundur's 4-Unit-13-bus System

In this section, the Kundur's 4-unit-13-bus system is used to validate the proposed strategies. For the convenience of analysis, the system is not divided into multiple interconnected subsystems. LFC in this case adopts the flat frequency control (FFC) mode. The area control error (ACE) is denoted by $ACE = \beta_a \omega_{COI}$. Supposing that an LAA occurs at the load bus 4, the defender uses all four units to counteract LAAs. The schematic of the Kundur's 4-unit-13-bus system under LAA is shown in Fig. 7.

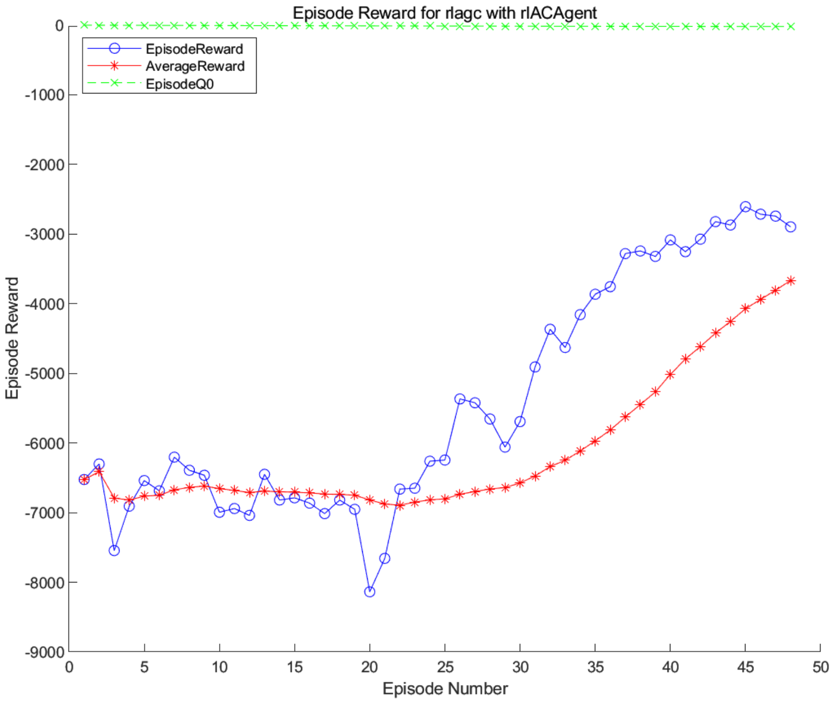


Figure 8: Moving episode rewards of AC-based RL for LAA $d = 5.p.u.$

5.1. Active Defense-Based Load Frequency Control Strategy

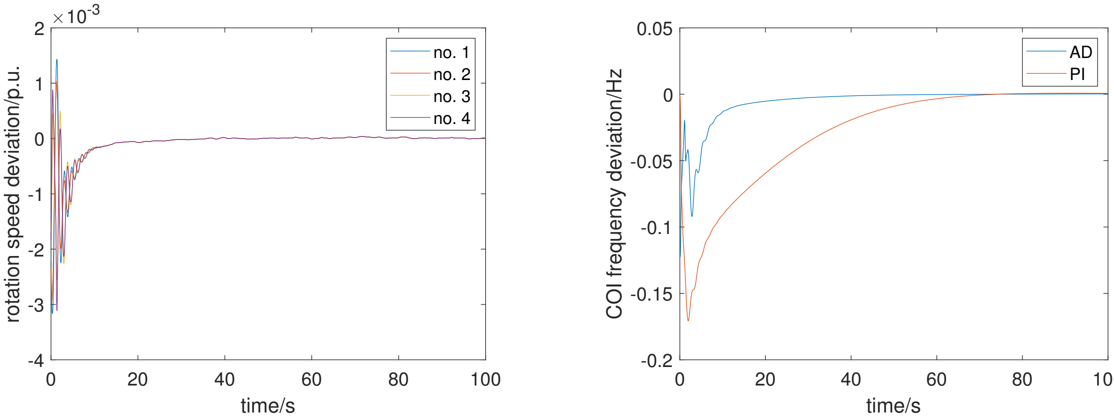
Based on Section 3, the model-free AD-based LFC strategy is simulated herein. Considering the uncertainty of LAAs, the LAA signal d is treated as a random variable conforming to some specific distributions. Uniform sampling can be used for larger number of samples. For the convenience of simulations with limited samples, it is supposed that the random d conforms to the normal distribution with $\mu = 5$ and $\sigma = 0.5$. After obtaining the samples $\mathcal{D} = \{d\}$ through Latin hypercube sampling, Algorithm 2 is executed for each d in $\mathcal{D}$. An exemplary training process for $d = 5.p.u.$ is shown in Fig. 8. As can be seen, both the episode and average reward eventually converge. Similarly, the 'optimal' off-policy can be obtained for all the 200 LAA scenarios represented by specific d values.

5.1.1. Active Defense Scenario Test

Supposing that during one round of LAA $d = 5.3p.u.$ (which is unknown to the defender), we first directly input the COI frequency under LAA into the regression network to obtain the estimation $\hat{d}$, then the optimal AD defense strategy is obtained by matching $\hat{d}$ with the member in the offline-trained strategy pool. By the aid of the trained network, after performing the matching and feedback compensation in Step 4 and 5 in Algorithm 1, the system frequency response is shown as the blue curve in Fig. 9. For comparison, the conventional PI-based feedback control (without considering the time delay) is also simulated and the dynamic response is shown as the red curve. As can be seen, even without the time delay, which is prevalent in the remote communication-based LFC, the dynamic COI frequency response under PI-based control is still inferior to the proposed AD strategy concerning the overshoot and settling time. It proves the superiority of the proposed AD strategies concerning both the stability and control velocity.

Also, as is known, when the attacker injects a major LAA signal, the abrupt excessive frequency deviation would occur and conventional feedback control schemes (e.g., PI) cannot be used, which conforms to the standard operating code in the emergency control. In this case, the proposed AD strategy, which behaves in the manner of the feedforward compensation by removing the correction module, can still be applicable to deal with major disturbances.

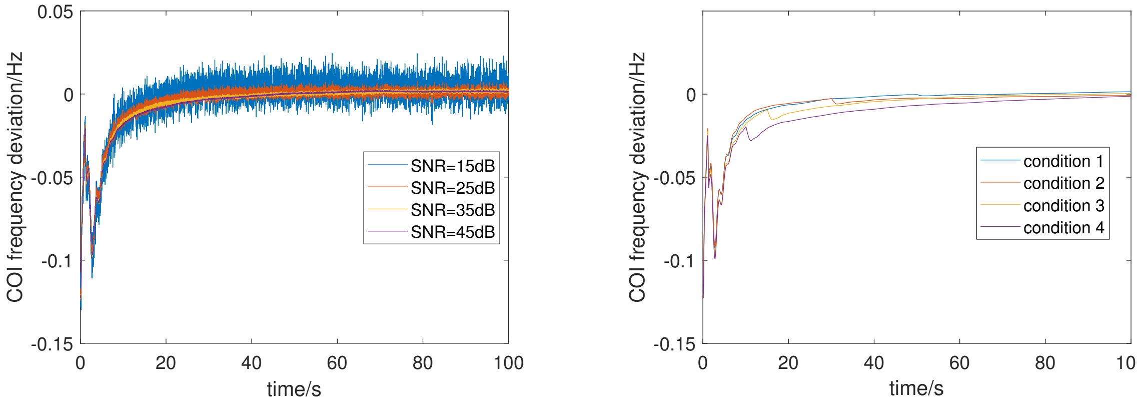
Noisy data might affect the learning performance by increasing irrelevant information and the learning model complexity. To validate the noise tolerance performance of the AD strategy, the frequency quality under different signal-to-noise ratios (SNRs) is demonstrated. The LAA signal is still set $d = 5.3p.u.$. The dynamic responses of COI frequency deviations under different SNRs in Fig. 10a show that the learning model is still effective at specific



(a) dynamic response of rotor speed deviation of the four units under the AD strategy

(b) dynamic response of COI frequency under AD and PI strategies

Figure 9: Dynamic responses of the Kundur's 4-unit-13-bus system in a AD scenario



(a) dynamic response of COI frequency under different SNRs

(b) dynamic response of COI frequency under different normal load variations

Figure 10: Robustness test of the AD strategy for the Kundur's 4-unit-13-bus system

noise levels. Besides the noise tolerance, robustness of the AD strategy under normal load variations is also tested. Supposing that an LAA occurs simultaneously with the normal load disturbance; and the disturbance conditions are: 1) $p_d = 0.01 p.u.$ at $t = 50s$; 2) $p_d = 0.05 p.u.$ at $t = 30s$; 3) $p_d = 0.1 p.u.$ at $t = 15s$; and 4) $p_d = 0.15 p.u.$ at $t = 10s$. The dynamic responses of COI frequency deviations in Fig. 10b show that the AD strategy is generally insensitive to the normal load variations with small magnitude.

5.2. Passive Defense-Based Load Frequency Control Strategy

Based on Section 4, the PD strategy is simulated herein. The LAA scenario is the same as Fig. 7, and the LAA signal is randomly selected among $(3, 7)$. The training process represented by the moving episode reward is shown in Fig. 11. As can be seen, the episode reward is significantly improved with the increase of episode numbers. It should be mentioned that due to the stochastic learning behavior, the absolute convergence cannot be achieved. Instead, the reward keeps changing at an acceptable level. After training the agent, the following PD scenario is tested.

5.2.1. Passive Defense Scenario Test

In this scenario, it is assumed that the LAA signal is $d = 5.3 p.u.$ (which is unknown to the defender). The 'optimal' PD defense strategy is obtained by directly inputting the COI frequency under this LAA into the offline-trained DQN

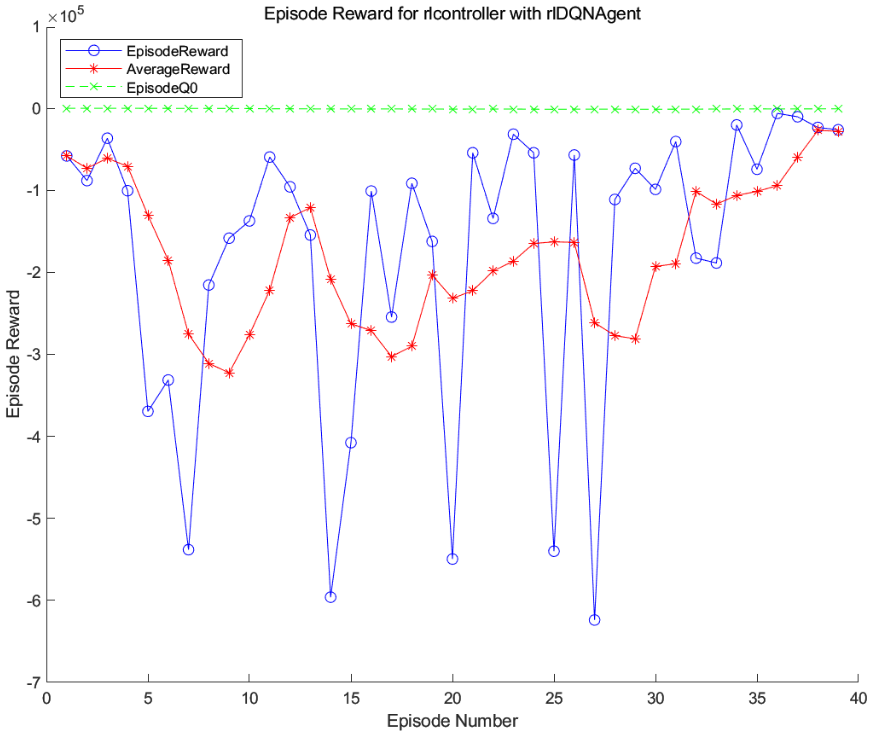
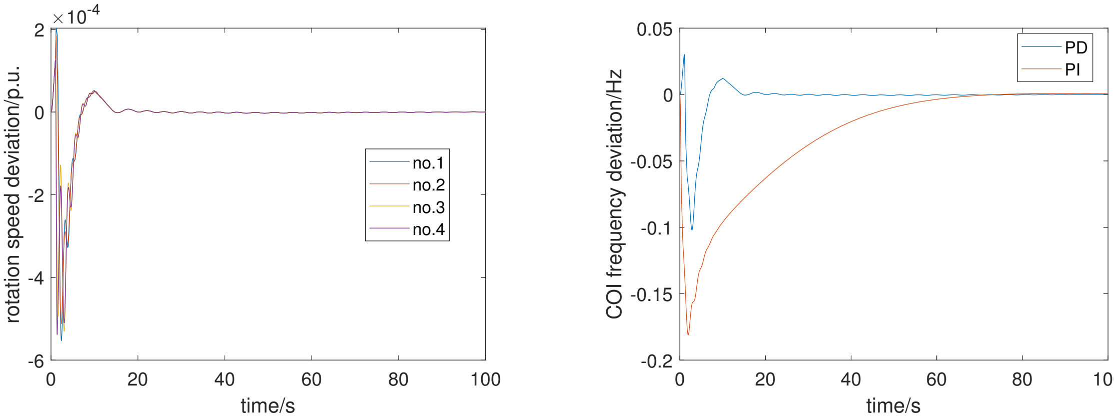


Figure 11: Moving episode rewards during the training process of DQN

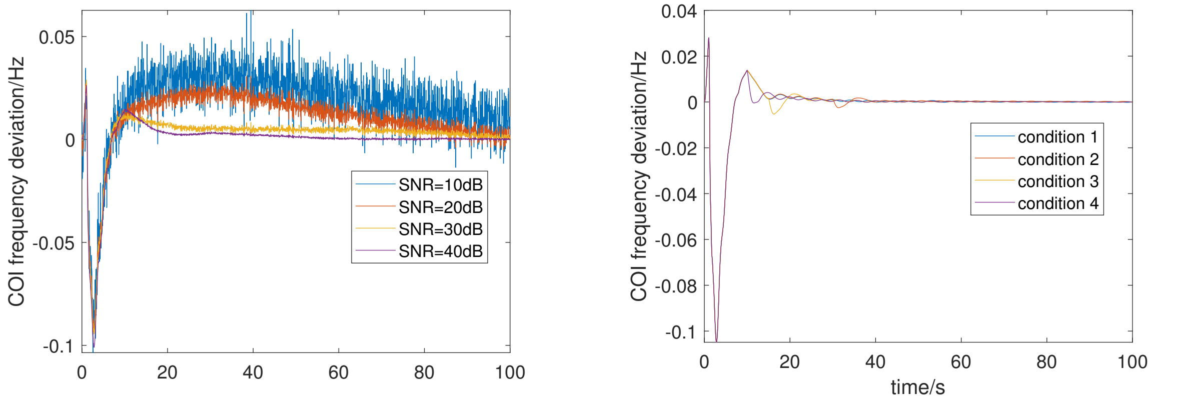


(a) dynamic response of rotor speed deviation of the four units under the PD strategy

(b) dynamic response of COI frequency under PD and PI strategies

Figure 12: Dynamic responses of the Kundur's 4-unit-13-bus system in a PD scenario

agent. By the aid of the trained DQN agent, the system frequency response is shown as the blue curve in Fig. 12. Also, it can be seen that the dynamic COI frequency response under the proposed PD strategy is significantly improved compared with PI concerning the overshoot and settling time. The main difference between the AD and PD strategies lies in the training cost of the RL agent. Since no deep neural network is required for AD, the training cost is low; nevertheless, more samples should be generated to enhance the robustness, and the overall cost might vary depending on the sample numbers. As for the PD strategies, the training cost concerning DQN would significantly increase with hyperparameters. Though the defender has no knowledge of d during the online operation process, no online estimation is required.



(a) dynamic response of COI frequency under different SNRs

(b) dynamic response of COI frequency under different normal load variations

Figure 13: Robustness test of the PD strategy for the Kundur's 4-unit-13-bus system

Robustness tests are also performed for the PD strategy in face of the noisy data and normal operating scenarios (normal load variations). The four normal operating conditions are the same as Fig. 10b. From Fig. 13a and 13b, it is learned that the passive strategy can also tolerate certain noise levels and normal load disturbances.

6. Case Studies for the IEEE 16-Unit-68-bus System

In this section, the proposed AD and PD strategies are simulated for the IEEE 16-unit-68-bus system. The system is treated as an 'aggregated' one for the convenience of analysis. The FFC mode is adopted herein and the ACE is denoted by $ACE = \beta_a \omega_{COI}$. And an LAA is supposed to occur at the load bus 52, 55, 56, and 67, the schematic is shown in Fig. 14. Also, it is assumed that only 4 (No. 13, 14, 15, 16) of the 16 units participate in the active defense. These four units have much larger capacity, and they can offer sufficient power support to attenuate the influence of LAAs. Also, if all of the 16 units participate in the regulation, the coordination among units with different dynamic characteristics and parameters might be considered to balance different performance indicators, which complicates the AD or PD strategy and is unnecessary.

6.1. Active Defense-Based Load Frequency Control Strategy

Based on Section 3, the AD strategy is simulated herein. As for the sample generation in Section 3.2.1, it is supposed that the LAA signal conforms to the normal distribution with $\mu = 7$ and $\sigma = 0.5$. After obtaining the samples $\mathcal{D} = \{d\}$ through Latin hypercube sampling, Algorithm 2 is executed for each d in $\mathcal{D}$. An exemplary training process for $d = 5.p.u.$ is shown in Fig. 15. As can be seen, both the episode and average reward finally converge to an acceptable value after probably 200 episodes of training. Similarly, the 'optimal' off-policy can be obtained for all the 200 LAA samples. As with Section 5.1, the regression network can be trained. After obtaining the optimal strategy pool and the trained network, the following AD scenario is tested.

6.1.1. Active Defense Scenario Test

Supposing that during one round of LAA $d = 5p.u.$ (which is unknown to the defender). By the aid of the trained network, after performing the matching and feedback compensation in Step 4 and 5 in Algorithm 1, the system frequency response is shown as the blue curve in Fig. 16. For comparison, the conventional PI-based feedback control is also simulated and dynamic response is shown as the red curve. As with Fig. 9, the dynamic COI frequency response under the proposed AD strategy is significantly improved compared with PI. It proves that the proposed feedforward compensation-based AD strategy can perform better than the conventional feedback ones no matter how large the system is.

To validate the noise tolerance performance of the AD strategy, the frequency control performance under different signal-to-noise ratios (SNRs) is demonstrated for the AD strategy. The dynamic responses of COI frequency deviations

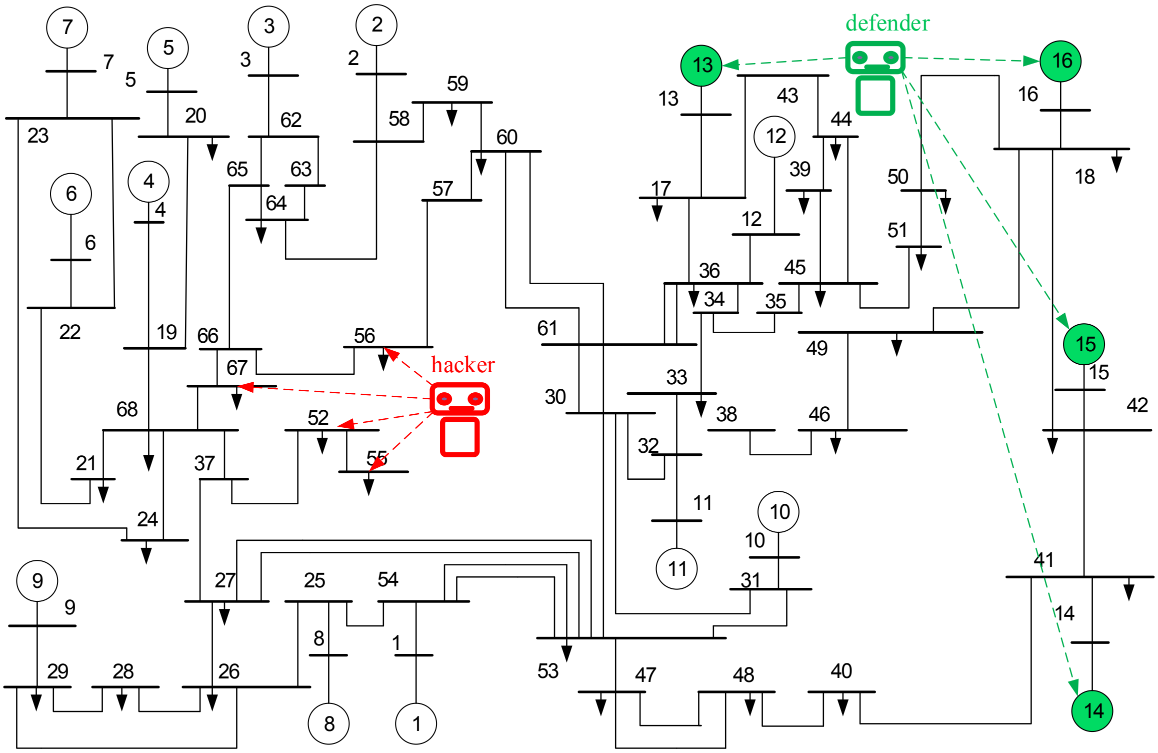


Figure 14: Load altering attacks on the IEEE 16-unit-68-bus system

under different SNRs in Fig. 17a show that the learning model is still effective at specific noise levels. Supposing that the normal load disturbance occurs at bus 60, and the four conditions are: 1) $p_d = 0.01p.u.$ at $t = 50s$; 2) $p_d = 0.05p.u.$ at $t = 30s$; 3) $p_d = 0.1p.u.$ at $t = 15s$; and 4) $p_d = 0.15p.u.$ at $t = 10s$. The dynamic responses of COI frequency deviations in Fig. 17b show that the AD strategy is generally insensitive to the normal load variations with small magnitude.

6.2. Passive Defense-Based Load Frequency Control Strategy

Based on Section 4, the model-free PD-based LFC strategy is simulated herein. The LAA scenario is the same as Fig. 7, and the LAA signal is randomly selected among $(3, 7)$. The training process represented by the episode reward is shown in Fig. 18. After training the agent, the following PD scenario is tested.

6.2.1. Passive Defense Scenario Test

In this scenario, it is assumed that the LAA signal is $d = 5.9p.u.$ (which is unknown to the defender). By the aid of the trained DQN agent, the system frequency response is shown as the blue curve in Fig. 19. Also, it can be seen that the dynamic COI frequency response under the proposed PD strategy is significantly improved compared with PI concerning the overshoot and settling time.

Robustness tests are also performed for the PD strategy in face of the noisy data and normal operating scenarios (normal load variations). The four load variation conditions are the same as Fig. 17b. From Fig. 20a and 20b, it is learned that the passive strategy can tolerate certain noise levels and normal load disturbances.

7. Conclusions

In this paper, model-free defense strategies for secondary frequency control are determined with the aid of reinforcement learning and deep neural network techniques. As can be seen from the simulation results, the proposed model-free AD and PD strategies can both handle LAAs with acceptable frequency control performances. In AD the defender would learn from a specific LAA scenario and actively optimize its defense policy for this scenario. In PD the

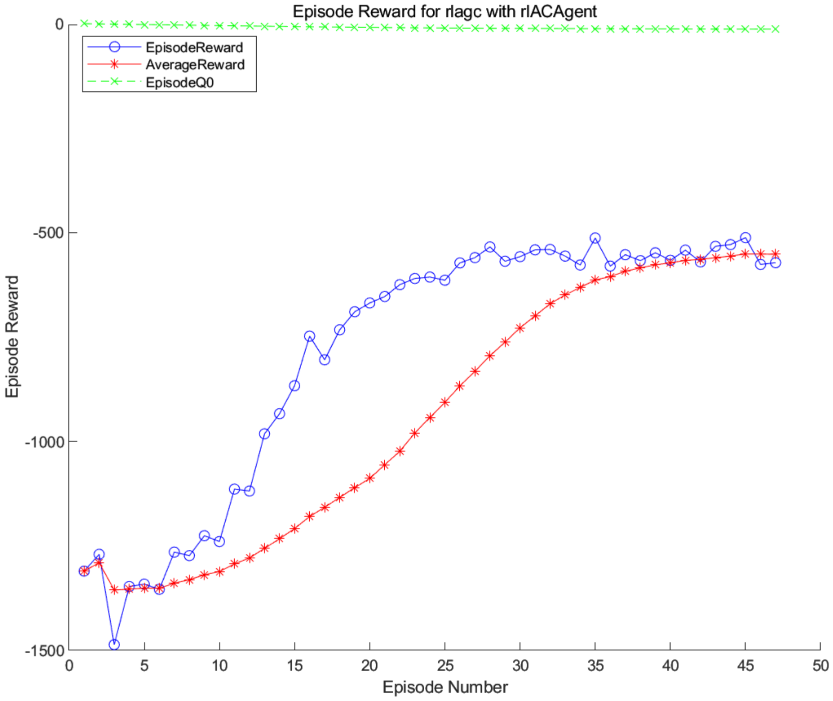
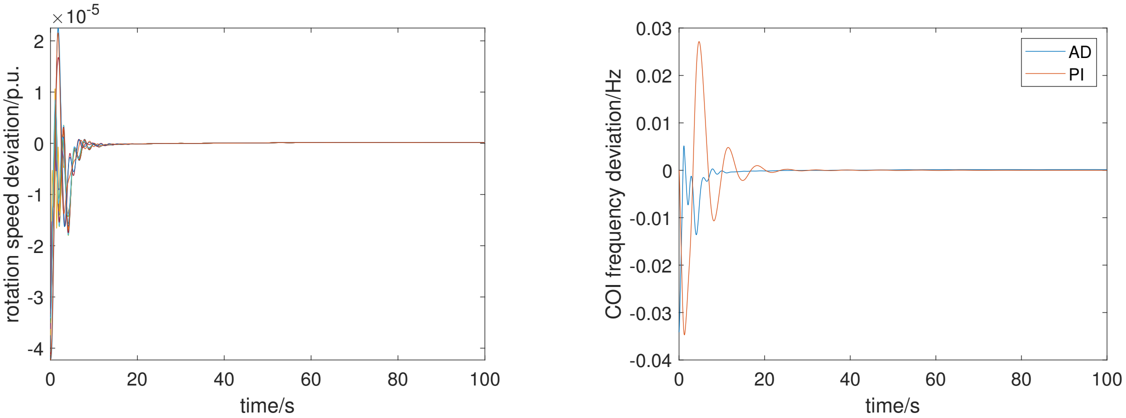


Figure 15: Moving episode rewards of AC-based RL for LAA $d = 5.p.u.$



(a) dynamic response of rotor speed deviation of the 16 units under the AD strategy

(b) dynamic response of COI frequency under AD and PI strategies

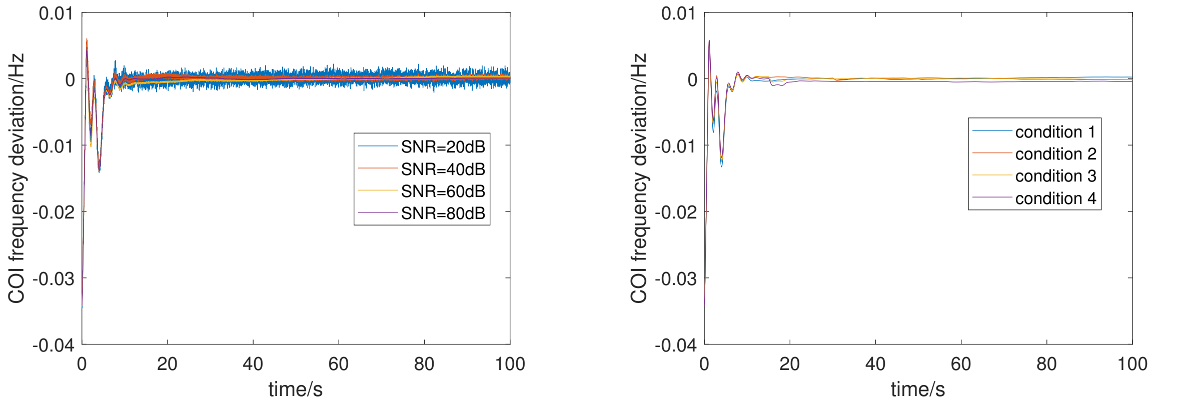
Figure 16: Dynamic responses of the IEEE 16-unit-68-bus system in a AD scenario

defender instead passively tolerates all potential LAA scenarios by enhancing the system redundancies. Both methods are applicable and the ultimate decision depends on whether the defender prioritizes the strong or weak initiative. To conclude, this paper presents a framework of model-free AD and PD strategies for electrical power system control under cyber attack. In future work, more types of cyber attacks will be studied under this proposed framework.

References

[1] M. Jin, R. Jain, C. Spanos, Q. Jia, Energy-cyber-physical systems, Applied Energy 256 (2019) 113939.

[2] X. Shang-Guan, Y. He, C. Zhang, L. Jiang, J. W. Spencer, M. Wu, Sampled-data based discrete and fast load frequency control for power systems with wind power, Applied Energy 259 (2020) 114–202.



(a) dynamic response of COI frequency under different SNRs

(b) dynamic response of COI frequency under different normal load variations

Figure 17: Robustness test of the AD strategy for the IEEE 16-unit-68-bus system

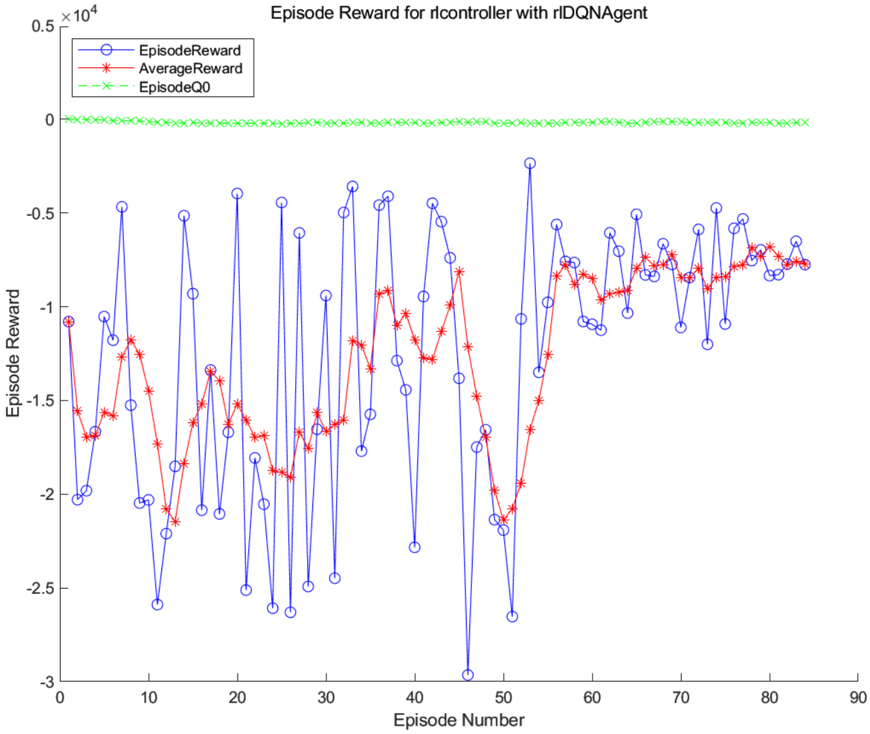


Figure 18: Moving episode reward during the training process of DQN

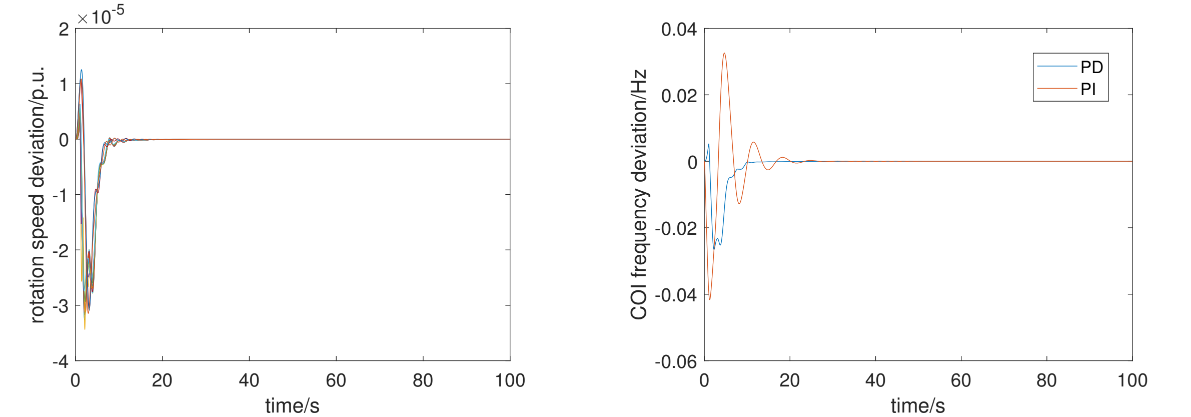
[3] S. Mishra, K. Anderson, B. Miller, K. Boyer, A. Warren, Microgrid resilience: A holistic approach for assessing threats, identifying vulnerabilities, and designing corresponding mitigation strategies, Applied Energy 264 (2020) 114726.

[4] M. Cui, J. Wang, B. Chen, Flexible machine learning-based cyberattack detection using spatiotemporal patterns for distribution systems, IEEE Transactions on Smart Grid 11 (2020) 1805–1808.

[5] X. Luo, X. Wang, M. Zhang, X. Guan, Distributed detection and isolation of bias injection attack in smart energy grid via interval observer, Applied Energy 256 (2019) 113703.

[6] C. Chen, K. Zhang, K. Yuan, L. Zhu, M. Qian, Novel detection scheme design considering cyber attacks on load frequency control, IEEE Transactions on Industrial Informatics 14 (2017) 1932–1941.

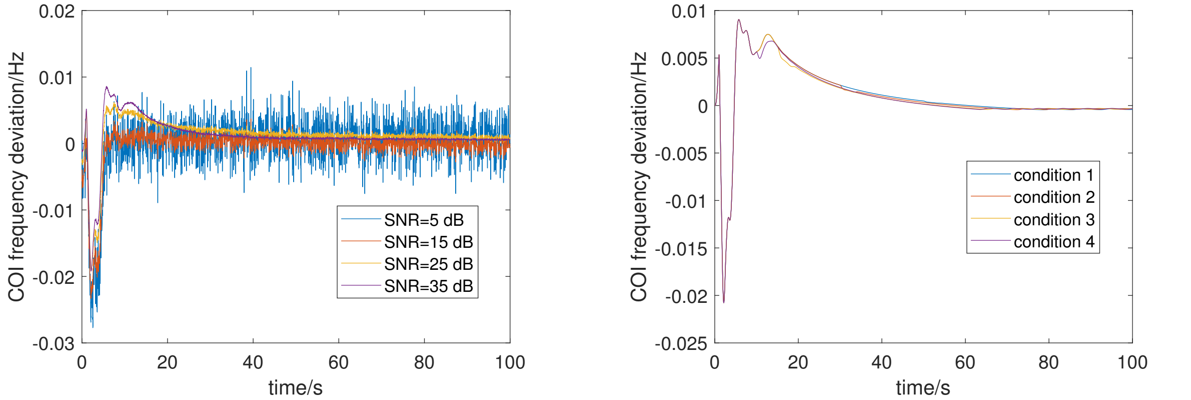
[7] P. Li, Y. Liu, H. Xin, X. Jiang, A robust distributed economic dispatch strategy of virtual power plant under cyber-attacks, IEEE Transactions



(a) dynamic response of rotor speed deviation of the 16 units under the PD strategy

(b) dynamic response of COI frequency under PD and PI strategies

Figure 19: Dynamic responses of the IEEE 16-unit-68-bus system in a PD scenario



(a) dynamic response of COI frequency under different SNRs

(b) dynamic response of COI frequency under different normal load variations

Figure 20: Robustness test of the PD strategy for the IEEE 16-unit-68-bus system

on Industrial Informatics 14 (2018) 4343–4352.

[8] Y. Tan, Y. Li, Y. Cao, M. Shahidehpour, Y. Cai, Severe cyber attack for maximizing the total loadings of large-scale attacked branches, IEEE Transactions on Smart Grid 9 (2018) 6998–7000.

[9] R. Deng, G. Xiao, R. Lu, H. Liang, A. V. Vasilakos, False data injection on state estimation in power systems attacks, impacts, and defense a survey, IEEE Transactions on Industrial Informatics 13 (2016) 411–423.

[10] Q. Yang, D. An, R. Min, W. Yu, X. Yang, W. Zhao, On optimal PMU placement-based defense against data integrity attacks in smart grid, IEEE Transactions on Information Forensics and Security 12 (2017) 1735–1750.

[11] S. K. Singh, K. Khanna, R. Bose, B. K. Panigrahi, A. Joshi, Joint-transformation-based detection of false data injection attacks in smart grid, IEEE Transactions on Industrial Informatics 14 (2017) 89–97.

[12] X. Liu, Z. Li, Z. Shuai, Y. Wen, Cyber attacks against the economic operation of power systems: A fast solution, IEEE Transactions on Smart Grid 8 (2016) 1023–1025.

[13] C. Zhao, J. He, P. Cheng, J. Chen, Analysis of consensus-based distributed economic dispatch under stealthy attacks, IEEE Transactions on Industrial Electronics 64 (2016) 5107–5117.

[14] X. Liu, Z. Bao, D. Lu, Z. Li, Modeling of local false data injection attacks with reduced network information, IEEE Transactions on Smart Grid 6 (2015) 1686–1696.

[15] Y. Xiang, Z. Ding, Y. Zhang, L. Wang, Power system reliability evaluation considering load redistribution attacks, IEEE Transactions on Smart Grid 8 (2016) 889–901.

[16] Y. Liu, S. Gao, J. Shi, X. Wei, Z. Han, Sequential-mining-based vulnerable branches identification for the transmission network under continuous load redistribution attacks, IEEE Transactions on Smart Grid (2020).

[17] Z. Li, M. Shahidehpour, A. Alabdulwahab, A. Abusorrah, Analyzing locally coordinated cyber-physical attacks for undetectable line outages, IEEE Transactions on Smart Grid 9 (2018) 35–47.

[18] C. Chen, M. Cui, X. Wang, K. Zhang, S. Yin, An investigation of coordinated attack on load frequency control, IEEE Access 6 (2018) 30414–30423.

[19] Q. Wang, W. Tai, Y. Tang, M. Ni, S. You, A two-layer game theoretical attack-defense model for a false data injection attack against power systems, International Journal of Electrical Power & Energy Systems 104 (2019) 169–177.

[20] K. Lai, M. Illindala, K. Subramaniam, A tri-level optimization model to mitigate coordinated attacks on electric power systems in a cyber-physical environment, Applied Energy 235 (2019) 204–218.

[21] A. Latif, S. S. Hussain, D. C. Das, T. S. Ustun, State-of-the-art of controllers and soft computing techniques for regulated load frequency management of single/multi-area traditional and renewable energy based power systems, Applied Energy 266 (2020) 114858.

[22] S. Amini, F. Pasqualetti, H. Mohsenian-Rad, Dynamic load altering attacks against power system stability: Attack models and protection schemes, IEEE Transactions on Smart Grid 9 (2018) 2862–2872.

[23] P. M. Esfahani, M. Vrakopoulou, K. Margellos, J. Lygeros, G. Andersson, Cyber attack in a two-area power system: Impact identification using reachability, in: Proceedings of the 2010 American control conference, IEEE, 2010, pp. 962–967.

[24] C. Chen, Y. Chen, K. Zhang, M. Ni, S. Wang, R. Liang, System redundancy enhancement of secondary frequency control under latency attacks, IEEE Transactions on Smart Grid (2020).

[25] M. Vrakopoulou, P. M. Esfahani, K. Margellos, J. Lygeros, G. Andersson, Cyber-attacks in the automatic generation control, in: Cyber Physical Systems Approach to Smart Electric Power Grid, Springer, 2015, pp. 303–328.

[26] R. Tan, H. H. Nguyen, E. Y. Foo, D. K. Yau, Z. Kalbarczyk, R. K. Iyer, Modeling and mitigating impact of false data injection attacks on automatic generation control, IEEE Transactions on Information Forensics and Security 12 (2017) 1609–1624.

[27] S. Siddharth, G. Manimaran, Model-based attack detection and mitigation for automatic generation control, IEEE Transactions on Smart Grid 5 (2014) 580–591.

[28] Y. W. Law, T. Alpcan, M. Palaniswami, Security games for risk minimization in automatic generation control, IEEE Transactions on Power Systems 30 (2014) 223–232.

[29] M. Khalaf, A. Youssef, E. El-Saadany, Joint detection and mitigation of false data injection attacks in agc systems, IEEE Transactions on Smart Grid (2018).

[30] A. Ameli, A. Hooshyar, E. F. El-Saadany, A. M. Youssef, Attack detection and identification for automatic generation control systems, IEEE Transactions on Power Systems 33 (2018) 4760–4774.

[31] A. R. Sayed, C. Wang, T. Bi, Resilient operational strategies for power systems considering the interactions with natural gas systems, Applied Energy 241 (2019) 548–566.

[32] S. Fan, G. He, X. Zhou, M. Cui, Online optimization for networked distributed energy resources with time-coupling constraints, IEEE Transactions on Smart Grid (2020).

[33] A. A. Saad, S. Faddel, O. Mohammed, A secured distributed control system for future interconnected smart grids, Applied Energy 243 (2019) 57–70.

[34] C. Wei, Y. Liao, W. Xi, Z. Yin, J. Luo, Event-driven adaptive fault-tolerant tracking control for uncertain mechanical systems with application to flexible spacecraft, Journal of Vibration and Control (2020).

[35] P. Wu, J. Partridge, R. Bucknall, Cost-effective reinforcement learning energy management for plug-in hybrid fuel cell and battery ships, Applied Energy 275 (2020) 115258.

[36] W. Zhang, J. Wang, Y. Liu, G. Gao, S. Liang, H. Ma, Reinforcement learning-based intelligent energy management architecture for hybrid construction machinery, Applied Energy 275 (2020) 115401.

[37] Y. Li, H. He, A. Khajepour, H. Wang, J. Peng, Energy management for a power-split hybrid electric bus via deep reinforcement learning with terrain information, Applied Energy 255 (2019) 113762.

[38] A.-H. Mohsenian-Rad, A. Leon-Garcia, Distributed internet-based load altering attacks against smart power grids, IEEE Transactions on Smart Grid 2 (2011) 667–674.

[39] M. Fliess, C. Join, Model-free control, International Journal of Control 86 (2013) 2228–2252.

[40] J. Jiang, X. Yu, Fault-tolerant control systems: A comparative study between active and passive approaches, Annual Reviews in Control 36 (2012) 60–72.

[41] R. Isermann, Model-based fault-detection and diagnosis–status and applications, Annual Reviews in Control 29 (2005) 71–85.

[42] C. Chen, M. Cui, F. F. Li, S. Yin, X. Wang, Model-free emergency frequency control based on reinforcement learning, IEEE Transactions on Industrial Informatics (2020).

[43] F. Shahid, A. Zameer, A. Mehmood, M. A. Z. Raja, A novel wavenets long short term memory paradigm for wind power prediction, Applied Energy 269 (2020) 115098.