Article

A Sparse Neural Network Based Control Structure Optimization Game under DoS Attacks for DES Frequency Regulation of Power Grid

Jian Sun ¹ , Guanqiu Qi ^{2,*} and Zhiqin Zhu ^{3,*}¹ School of Electronic and Information Engineering, Southwest University, Chongqing 400715, China; cq_jsun@163.com² Computer Information Systems Department, Buffalo State College, Buffalo, NY 14222, USA³ College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

* Correspondence: qig@buffalostate.edu (G.Q.); zhuzq@cqupt.edu.cn (Z.Z.)

Received: 22 April 2019; Accepted: 26 May 2019; Published: 30 May 2019

Abstract: With the rapid growth of distributed energy sources, power grid has become a flexible and complex networked control system. However, it increases the chances of being a denial-of-service attack, which degrades the performance of the power grid, even causing cascading failures. To mitigate negative effects from denial-of-service attack and enhance the reliability of the power grid, we propose a networked control system structure based optimization scheme that is derived from a Stackelberg game model for the frequency regulation of a power grid with distributed energy sources. In the proposed game model, both denial-of-service attacker and control system designer as a defender are considered without using any analytical model. For defenders, we propose a sparse neural network based DES control and system structure design scheme. The neural network is used to approximate the desired control output and reinforce signals for the improvements of short- and long-term performance. It also introduces the sparse regulation of column grouping in the neural network learning process to explore the structure of control system that involves the placement of sensor, distributed energy sources actuator, and communication topology. For denial-of-service attackers, the related attack constraints and attack rewards are established. The solution of game equilibrium is considered as an optimal solution for both denial-of-service attack strategy and control structure. An offline optimization algorithm is proposed to solve the game equilibrium. The effectiveness of proposed scheme is verified by two cases, which illustrate the optimal solutions of both control structure and denial-of-service attack strategy.

Keywords: DoS attack; reinforcement learning; neural network; distributed energy resources; power grid; smart grid

1. Introduction

With the development of network techniques, electricity supply via the modern power grid is increasingly depending on networked control systems (NCSs). In the power grid and communication integrated network, the efficiency and reliability of power grid are gradually enhanced [1,2]. However, new network control techniques generate related vulnerabilities in the control system of power grids. As the connection of virtual and physical worlds, cyber attacks against NCSs can render the large disturbances of power grid that have been confirmed during the past few years [3].

As a common type of cyber attack, denial-of-service (DoS) attack occupies communication resources to prohibit the transmission of measurement and control signals in NCSs [1]. Compared with deception attacks [4], DoS attacks not only require little prior knowledge about NCSs, but also

destroy control operations in real time [5]. In particular, blocking real-time control signals can cause the instability of the power grid [6]. A lot of security control approaches, such as stochastic time delay system, triggering strategy, and game theory, have been applied to the prevention of DoS attacks in power grid.

Stochastic time delay system: In a stochastic time delay system, DoS is modeled as a stochastic process with a delay in the signal. Subject to intermittent DoS attacks, An et al. proposed a decentralized solution of adaptive output feedback control for a power grid [7]. A switching-type state estimator is presented to estimate the state of power grid by using discontinuous output measurement. Sun et al. modeled DoS attacks as a Markov process that converted DoS attacks to stochastic noises of NCSs first [8]. Then, a resilient control model was used to process the converted noises.

Triggering strategy: Event-triggering time-sequence of control signal is adopted to reduce communication costs in the system. The triggering time sequence is able to defend against DoS attacks. Peng et al. proposed a resilient event-triggering based frequency control method for power system control in energy-limited DoS attacks [1]. The proposed event-triggered communication scheme can tolerate a certain degree of data loss in the open communication induced by energy-limited DoS attacks. Hu et al. designed a periodically resilient event-triggering communication scheme to identify DoS attacks initiated by power-constrained pulse-width-modulated jammers [9].

Game theory: In game theory based cyber security model, a game model is constructed for both DoS attacker and defender to obtain a Nash equilibrium as an optimal solution. Li et al. modeled the interactions between the transmission power of sensors and the interference power of DoS attacker by a signal-to-interference-and-noise-ratio (SINR) based network [10]. A modified Nash Q-learning algorithm was proposed to analyze the related interactions as well. Yuan et al. analyzed the resilient control issues of NCS under DoS attack via a unified game approach [5]. Yuan et al. also built a multi-stage hierarchical game with a corresponding hierarchy of decisions that was implemented to achieve a resilient control system [11]. According to optimal control structures, optimal criteria were constructed for DoS attackers and cyber defenders. Ding et al. modeled the remote estimation under DoS attacks by using the strategy of zero-sum stochastic games and presented a monotone structure to solve the proposed model optimistically [12]. At the same time, Ding et al. also formulated the decision-making process of a target channel as a two-player zero-sum stochastic game framework and proposed a Nash Q-learning algorithm to obtain the optimal strategy [13]. Zhu et al. built a hybrid game-theoretic framework to improve the robustness of power system security [14]. Since game theory based control strategy can process strategic interactions among multiple decision makers, it is often used in large-scale system security control.

Despite the game theory based method being able to be efficient against conventional DoS threats, only using control methods to defense DoS attacks is not sufficient to counter potential and increasingly sophisticated attacks. In large-scale control systems, structure factors, such as distributed energy sources (DES) placements, sensors, and communication link topologies, can critically affect the performance in DoS attacks. Existing work of energy storage systems [15], DES placement [16], sensor scheduling [17] and coverage aims at minimizing the electrical and computational costs [18]. Upon existing research in the optimization of control system structure for cyber security, there are several critical challenges. One critical challenge is the transient and dynamic issues of DES and cyber resources allocation. Most existing research focused on the characteristics of steady state [19]. They ignored the transient and dynamic issues in control systems [20]. Another critical challenge is that most existing algorithms used a precisely analytical model as a premise [21]. However, most large-scale power systems are unable to be precisely modeled. In this case, a model free method is required for control-system-structure optimization.

With the consideration of DoS [22,23] attacks and the control-system-structure optimization problem, we propose a sparse neural network based NCS optimization method that is derived from the Stackelberg game model for the frequency regulation of DES in power systems. A neural controller is trained by reinforcement learning of offline simulation and online processes to improve the system

performance. The system performance is also improved under limited cyber resources by optimizing the control structure, which involves the placement of DES, RTU sensors and communication topology. It is optimized by imposing a group sparse regulation on neural network weights. The Stackelberg game model is used, which derives a minimax optimization of system performance, so that the optimized control structure is robust to DoS attacks under the consideration of worst case attacks. The structure consists of the placement of DES, RTU sensors and communication topologies. The contributions of this paper are summarized as follows:

1. Sparse neural network based reinforcement learning is proposed to improve the frequency regulation of DES in control systems without using a power system analytical model, which involves adaptiveness, performance, and structure.
2. The Stackelberg game model is used to derive the optimal control scheme and structure, so that the proposed frequency regulation system is robust to the worst case of DoS attacks. In addition, the reliability of proposed frequency regulation system is enhanced.

The remainder of this paper is organized as follows: the system model and related problems are formulated in Section 2, which introduces the power system and DoS attack model, and describes the frequency regulation as well; Section 3 elaborates control structure and control law design by sparse neural network based reinforcement learning; Stackelberg game model and the optimization scheme of control structure are derived under DoS attacks in Section 4; and Section 5 demonstrates the simulation results to verify the proposed algorithm.

2. Problem Formulation

The formulation of power grid and control objective will be introduced in this section. We consider a multi-area system that is integrated with Distributed Energy resources. The control objective is to mitigate the frequency regulation. The DoS attack may degrade the control system performance and even lead to failures. Thus, a design problem involving controller design and structure design under the cyber attack is also introduced.

2.1. Power Grid Frequency Dynamic Model

We consider the interconnected multi-area power system. Each one of n areas is connected to each other by tie-line (also called transmission line). As shown in Figure 1, each area equips a turbine generator and DES, such as wind power, solar power, battery, etc. [20].

It also contains a load frequency controller (LFC) and tie-line bias controller (TBC) for frequency synchronization. Even if the synchronization measures regulate the frequency, an auxiliary control offered by DES may be necessary to enhance the system performance, when the power system encounters a severe disturbance, such as system fault, and sudden large load drop. Considering the auxiliary control, the dynamic model of area i can be formulated as a discrete linear difference equation [24]:

$$x_i(k+1) = A_i x_i(k) + B_i u_i(k) + \sum_{j \in N_p(i)} B_{ji} x_j(k) + E_i w_i(k), \quad (1)$$

where $i \in N^+ \cap [1, n]$, $x_i = \begin{bmatrix} \Delta f_i & \Delta P_{mi} & \Delta P_{vi} & \Delta P_{tie-i} & \int ACE_i \end{bmatrix}^T$ is area state. Δf_i is the deviation related to synchronized frequency; ΔP_{mi} is the mechanical power deviation of generator; ΔP_{vi} is the valve position deviation of turbine; ΔP_{tie-i} is the deviation of tie-line power injection from other physical neighbored areas; ACE_i is the ACE signal of area i and $ACE_i = \alpha_i \Delta f_i + \Delta P_{tie-i}$. u_i as the auxiliary control output of DES for frequency regulation is the sum of all the powers generated from power-electronic interfaced DES; w_i is the disturbance caused by model error or other time-varying factors; A_i is the system transition matrix; B_i and B_{ji} are the gains of control effect and other physical neighbored systems; E_i is the disturbance gain. $N_p(i)$ denotes the physical neighbored areas of area i .

In this linear model, loads are assumed to be constant because the variation of loads is slow relative to the dynamic frequency regulation. Therefore, A_i , B_i , and B_{ji} can be modeled in time invariant [24]. The system is linear time invariant (LTI) as well as an NCS. The DES controller of area i for frequency regulation are written as

$$u_i = \varphi_i(x_i, x_{N_c(i)}), \quad (2)$$

where the time stage k is neglected. $N_c(i)$ denotes the cyber connected areas of area i . The controller calculates DES control outputs by the received state x_i and x_j as well as $j \in N_c(i)$ from local and remote areas, respectively. The control objective is to mitigate the frequency deviation and reduce the overall costs defined in Section 3, which consider the least quadratic of state x_i and the control output u_i .

Figure 1. Area structure.

2.2. DoS Attack Model

For the previously mentioned NCS, we consider attacker launch attacks, when the power system requires the emergency auxiliary control offered by DES. DoS attack blocks communication channels to degrade the control performance, even causing system failures. The blocks of communication channels probably result in the absence of some remote states x_j and $j \in N_c(i)$ of the current time stage k [25]. When controller cannot obtain the remote state from cyber connected area j , “zero-control” strategy is applied to assigning $x_j = 0$. u_i becomes

$$u_i = \varphi[\{\alpha_{ji}x_j | j \in N_c(i) \cup i\}], \quad (3)$$

where $\alpha_{ji} \in \{0, 1\}$ is distributed according to Bernoulli distribution, which is

$$\begin{cases} P(\alpha_{ji}(k) = 0) = \delta_{ji}, \\ P(\alpha_{ji}(k) = 1) = 1 - \delta_{ji}, \end{cases} \quad (4)$$

where $\delta_{ji} \in [0, 1]$ is the probability of packet drop in the communication from area j to i , $k \in N^+$. The probability δ_{ji} increases the intensity of DoS attack. DoS attacker has the limited cyber resources, which constrains the intensity of DoS attack at a certain degree. The constraint in our model is subject to

$$\sum_{i=1}^n \sum_{j \in N_c(i)} \delta_{ji} \leq C. \quad (5)$$

The probability δ_{ji} is limited to the intensity of DoS attack. The intensity of DoS attack is under the constraints of cyber cost, which is also limited. Therefore, the sum of probability δ_{ji} must be less than a maximal real number C .

2.3. Control, Structure Design and Optimization Problem

Even though the NCS of power grid suffers DoS attacks, one of the NCS objectives is to enhance the system performance by mitigating the frequency deviation Δf_i and minimizing the control cost. To achieve the control objective, the control law φ_i should be investigated. Moreover, the design of control structure and its optimization also need to be considered. Since the control structure associates the placement of sensors (in remote terminal unit (RTU)), actuators (in DES) and communication topology, these placements seriously affect the control performance in a large-scale multi-area power system. Generally, the placed resources (RTU, DES, etc.) are expensive.

In mathematical form, the control design gives a φ_i to achieve the system performance. We define Q shown in Equation (6) as the control cost to measure system performance.

$$\varphi^* = \arg \min_{\varphi} Q, \quad (6)$$

where “argmin” here also denotes the functional minimization; the control structure design aims to select cyber connected area set $N_c(i)$ for $i = 1, \dots, n$, which is

$$N_c^* = \arg \min_{N_c} Q. \quad (7)$$

Therefore, as an optimization, it searches a combination of φ_i and $N_c(i)$ for $N_c(i)$, $i = 1, \dots, n$ to minimize the control cost in the worst case of DoS attack with constraint Equation (5) as shown in Equation (8):

$$\left(\varphi^*, N_c^*, \delta^* \right) = \arg \max_{\delta} \min_{\varphi, N_c} Q, \quad (8)$$

where $\varphi = \begin{pmatrix} \varphi_1 & \dots & \varphi_n \end{pmatrix}$, $N_c = \begin{bmatrix} N_c(1) & \dots & N_c(n) \end{bmatrix}$, $\delta = \begin{bmatrix} \delta_{1N_c(1)} & \dots & \delta_{nN_c(n)} \end{bmatrix}$.

We assume the attacker knows the information of the designed power system. The attacker can figure out the minimal system performance, when the controller reaches its maximal performance. The details of optimization in a Stackelberg game are discussed in Section 4.

3. Control and Structure Design

3.1. Control Design by Reinforcement Learning

Before introducing control design, a cost function is defined to measure the control cost Q . Control performance usually involves the quadratic of state and control output. Thus, we define the control cost Q by a control strategy utility function used in our previous work [24]:

$$Q_i(k) = \min_{u_i(k+j), j \in [0, \infty]} \sum_{t=0}^{\infty} \alpha^{N-t} p_i(k+t), \quad (9)$$

where $0 < \alpha < 1$ is the discount rate, and N is a positive integer. p_i as a binary performance index is defined as:

$$p_i(k) = \begin{cases} 1, & a_1 \|x_i(k)\| + a_2 \|u_i(k)\| \leq c, \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

where $\|\cdot\|$ denotes 2-norm, a_1 and a_2 are the weights of state and control cost respectively, and c is the threshold of performance indication. If the current cost is in an allowed range (less than c), the binary performance index shown in Equation (10) is 1; otherwise, it is 0. Binary performance index makes the

strategy utility function limited. When time horizon is limited, even the system diverges, and it can avoid a numerical crash in the learning process.

To prevent the system diverge, we construct a desired control output u_{di} to apply a damping rate $L_i \in R^{5 \times 5}$, $\|L_i\| < 1$ to the system, so that the power grid frequency dynamic becomes

$$x_i(k+1) = L_i x_i(k) + \sum_{j \in N_p(i) \setminus N_c(i)} B_{ji} x_j(k) + E_i w_i(k). \quad (11)$$

If we approximate u_{di} by a neural network output $\hat{u}_{di}$, the neural network is defined as

$$\hat{u}_{di}(k) = M_a W_i \phi(k), \quad (12)$$

where $\phi \in R^{nl \times 1}$ is the radial basis of neural network calculated from state $x_1, x_2, \dots, x_n$, $W_i \in R^{nl \times nl}$ is the trainable weight, and $M_a \in R^{1 \times nl}$ is a given constant matrix. As shown in Figure 2, the proposed neural network structure is based on a radial basis function (RBF). An RBF-based neural network can identify nonlinear dynamical systems [26]. l is the dimension of radius basis for one area. Under the control of $\hat{u}_{di} = u_i$, the system of area i becomes

$$x_i(k+1) = L_i x_i(k) + B_i [\hat{u}_{di}(k) - u_{di}(k)] + \sum_{j \in N_p(i) \setminus N_c(i)} B_{ji} x_j(k) + E_i w_i(k). \quad (13)$$

According to the control theory of stability, if $B_i [\hat{u}_{di}(k) - u_{di}(k)] + E_i w_i(k)$ and $\sum_{j \in N_p(i) \setminus N_c(i)} B_{ji} x_j(k)$ in the right part of Equation (13) are bounded, the system is ultimately uniformly bound (UUB) [27]. We assume the disturbance $w_i(k)$ is bounded as $\|w_i(k)\| \leq \varepsilon$, where ε is a small real number, so learning W_i should involve the approximation of u_{di} .

The control cost involving system performance is also approximated by a neural network as shown in Equation (14):

$$\hat{Q}_i(k) = M_c W_i \phi(k), \quad (14)$$

where $M_c \in R^{1 \times nl}$ is a given constant matrix.

Figure 2. Neural network structure.

We train the neural network weight W_i to approximate the desired control output u_{di} as well as the performance measurement for the improvement of system performance. Therefore, we define a loss function for training as

$$V_i(k) = \frac{1}{2} \|B_i \hat{u}_{di}(k) - B_i u_{di}(k)\|^2 + \frac{1}{2} \left\| \alpha^{N+1} p_i(k) + \alpha^{-1} \hat{Q}_i(k+1) - \hat{Q}_i(k) \right\|^2 + \frac{1}{2} \|\hat{Q}_i(k)\|^2. \quad (15)$$

The first term denotes the approximation error of the desired control output, the second term denotes the approximation error of system performance, and the third term is used to improve the long-term system performance. As shown in Equation (16), the desired control output u_{di} and system performance Q_i can be expressed by the neural network:

$$\begin{cases} u_{di}(k) = M_a W_i^* \phi(k) + v_{ai}(k), \\ Q_i(k) = M_c W_i^* \phi(k) + v_{ci}(k), \end{cases} \quad (16)$$

where $\|v_{ai}(k)\| \leq v$ and $\|v_{ci}(k)\| \leq v$ are the optimal approximation errors, v is a small number, and W_i^* is the optimal approximation of a constant matrix. We derive the defined loss function $V_i(k)$ subject to $W_i(k)$ and neglect v_{ai} and v_{ci} , so the converted online iteration formula is shown as

$$\begin{cases} W_i(k+1) = W_i(k) - \beta_i \frac{\partial L_i(k)}{\partial W_i(k)}, \\ \frac{\partial L_i(k)}{\partial W_i(k)} = \{M_a^T [x_i(k+1) - L_i x_i(k)] + M_c^T [\alpha^{N+1} p_i(k) + \alpha^{-1} \hat{Q}_i(k+1)]\} \phi(k)^T. \end{cases} \quad (17)$$

3.2. Structure Design by Sparse Neural Networks

The structure design involves the placement of sensor, actuator, and communication topology. It is formulated by the neural network weight W_i . The weight matrix is separated into column groups, so we can define the radius basis ϕ as

$$\phi = \begin{pmatrix} \phi_1^T & \phi_2^T & \cdots & \phi_n^T \end{pmatrix}^T, \quad (18)$$

where $\phi_i = \begin{pmatrix} e^{-\frac{\|x_i - x_{c1}\|^2}{\sigma^2}} & e^{-\frac{\|x_i - x_{c2}\|^2}{\sigma^2}} & \cdots & e^{-\frac{\|x_i - x_{cl}\|^2}{\sigma^2}} \end{pmatrix}^T$, x_{cj} is a given radius center, $j \in N^+ \cap [1, l]$, and σ is a given radius width. According to the definition of ϕ , we know that the neural weight is separated into n column groups. Each group that has l columns corresponds to the radius basis ϕ_i of one area. If the weight matrix W_i is group sparse [28] with respect to these columns, the structure of control system can be figured out. For instance, if the column group j of W_i corresponding to the gain of ϕ_i is zero, the controller of area i does not need the information from area j , and the communication channel between area j and i is also not necessary. If all the controllers do not require the information from area j , any sensor is not necessary for area j . If $W_i = 0$, it means that the actuator in area i is not required. The key is to force W_i to be column group sparse. Thus, a group sparse regulation term is added to the loss function. Thus, the loss function becomes

$$V_i(k) = \frac{1}{2} \left[\|B_i \hat{u}_i(k) - B_i u_{di}(k)\|^2 + \|\hat{Q}_i(k)\|^2 \right] + \frac{1}{2} \|\alpha^{N+1} p_i(k) + \alpha^{-1} \hat{Q}_i(k+1) - \hat{Q}_i(k)\|^2 + \gamma_i \sum_{j=1}^n \left\| \text{vec} \left(W_{iG(j)} \right) \right\|, \quad (19)$$

where $W_{iG(j)}$ denotes the column group j or the gain of radius basis ϕ_j . γ_i is the regulation weight. The sparse regulated learning iteration is demonstrated in Equation (20), which can be also derived by Equation (17):

$$\begin{cases} W_i(k+1) = W_i(k) - \beta_i \frac{\partial V_i(k)}{\partial W_i(k)} \\ \frac{\partial V_i(k)}{\partial W_i(k)} = \{M_a^T [x_i(k+1) - L_i x_i(k)] + M_c^T [\alpha^{N+1} p_i(k) + \alpha^{-1} \hat{Q}_i(k+1)]\} \phi(k)^T \\ \quad + \gamma_i W_i(k) D_i(k), \end{cases} \quad (20)$$

where $D_i = \text{diag} \left[\mathbf{1} \left\| \text{vec}(W_{iG(1)}) \right\|^{-1} \quad \mathbf{1} \left\| \text{vec}(W_{iG(2)}) \right\|^{-1} \quad \cdots \quad \mathbf{1} \left\| \text{vec}(W_{iG(n)}) \right\|^{-1} \right]$, and $\mathbf{1} = [1, 1, \dots, 1] \in \mathbb{R}^l$. W_{ij} means the j th weight block in W_i , which involves ϕ_j . Thus, we have

$$\hat{u}_{di} = W_{i1}\phi_1 + W_{i2}\phi_2 + \dots + W_{in}\phi_n. \quad (21)$$

The stability analysis of iteration in Equation (20) can be taken in a similar way used in [24]. We know that if β_i and γ_i are small enough, the iteration in Equation (20) is stable and the errors in results are limited.

4. Structure Optimization under DoS Attacks

In this section, the control system structure design is solved by a structure optimization problem. Considering DoS attack, a Stackelberg game is formulated in the structure optimization. In addition, the optimization algorithm is also presented.

4.1. Stackelberg Game Formulation

As shown in Equation (20), the iteration searches a system structure and control scheme in simulation, and then a DoS attacker observes the system and takes attack actions. It is a leader–follower sequence, therefore a Stackelberg game model is proposed to obtain the optimal solution of control structure optimization and DoS attacks.

There are two actors in the proposed model. One is a defender that is the system designer, and the other one is an attacker that is a DoS attacker. The defender's action is $N_c = \{N_c(1), N_c(2), \dots, N_c(n)\}$. The reward function of defender is defined as

$$r(N_c, \delta) = \sum_i^n Q_i|_{N_c(i), \delta}. \quad (22)$$

The attacker's action is $\delta = \left\{ \delta_{N_c(1)1} \quad \delta_{N_c(2)2} \quad \cdots \quad \delta_{N_c(n)n} \right\}$. The reward of attacker is $-r(N_c, \delta)$. Therefore, it is a zero-sum game. The structure optimization becomes a min-max optimization problem as follows:

$$(N_c^*, \delta^*) = \arg \min_{N_c} \max_{\delta} r(N_c, \delta), \quad (23)$$

where N_c^* and δ^* are the equilibrium of game model, which can also be treated as the optimal solution for both defender and attacker. Thus, it can obtain $r(N_c, \delta^*) \leq r(N_c^*, \delta^*) \leq r(N_c^*, \delta)$.

4.2. Structure Optimization under Dos Attacks

For a given DoS attack δ , we know that the optimal structure and control law can be figured out by the iteration of Equation (20): The iteration derives as follows.

$$N_c^*|_{\delta} = \arg \max_{N_c} r(N_c, \delta). \quad (24)$$

If $j \notin N_c^*(i)$, the iteration of Equation (20) usually obtains a small $\left\| \text{vec}(W_{ij}) \right\|$ instead of $\left\| \text{vec}(W_{ij}) \right\| = 0$ because it is an numerical method. Therefore, a threshold method is used in our algorithm to obtain $N_c^*|_{\delta}$. We can obtain the estimated $N_c^*|_{\delta}$ by Equation (25):

$$\hat{N}_c^*|_{\delta}(i) = \{j | \left\| \text{vec}(W_{ji}) \right\| \leq \rho_i\}, i \in \mathbf{N}^+ \cap [1, n], \quad (25)$$

where $\mathbf{N}^+$ is the integer set, and ρ_i is a given threshold that is a small positive number. The equilibrium of Stackelberg game can be reached by solving the optimization problem of Equation (23), or an optimization problem of minimization with constraints as follows:

$$\begin{aligned} \delta^* &= \arg \min_{\delta} r(N_c^* | \delta, \delta), \\ \text{s.t. } &\sum_{i=1}^n \sum_{j \in N_c(i)} \delta_{ji} \leq C, \delta_{ji} \in [0, 1] \quad j, i \in \mathbf{N}^+ \cap [1, n]. \end{aligned} \quad (26)$$

In summary, the algorithm of structure optimization under DoS attacks is described in Algorithm 1.

Algorithm 1 Algorithm of Structure Optimization under DoS Attacks

Input: : the radial basis dimension l , area number n , learning rate β_i and γ_i .

Output: : system structure N_c^* and optimal attack strategy δ^* .

- a. Initialize $M_a, M_c, W_i, i = 1, 2, \dots, n$ and $\delta(0)$, each element is subject to Gaussian distribution with small average value 0.01 and variance 0.14. Set $t = 0$ and maximal iteration number as MAXITER;
 - b. For given $\delta(t)$, find $\hat{N}_c^* |_{\delta(t)}$ by the iteration of Equation (20) and (25);
 - c. Measure the gradient of $r[N_c^* | \delta, \delta(t)]$ that is subject to δ ;
 - d. Update δ to obtain $\delta(t+1)$ based on the gradient obtained in step c, and handle the constraints of Equation (26) by Lagrange method [29] or barrier function method [30];
 - e. Check whether the updated variance is less than a threshold or t equals to the maximal iteration number MAXITER. If yes, end the algorithm and go to step f, otherwise $t = t + 1$ and return to step b;
 - f. Obtain $N_c^* = \hat{N}_c^* |_{\delta(t)}$ and $\delta^* = \delta(t)$.
-

The obtained N_c^* by Algorithm 1 is the optimized control structure of power system. The worst case of system performance with our controller is also known by Algorithm 1. However, we do not further consider the unstable control system caused by a large C [5] or the attacker has a lot of cyber resources to launch DoS attacks. These issues will be investigated in future.

5. Experiments and Analysis

This section illustrates two cases of IEEE 14 bus system and 24 bus system [31] to show the effectiveness and advantages of the proposed scheme. The scheme takes DoS attacks into account and considers the worst case of attacks under the constraints in the Stackelberg game model mentioned above. In the Stackelberg game model, the scheme uses the optimal structure design to enhance the system performance by reinforcement learning. We assume sub-system parameters are as follows:

$$A_i = \begin{pmatrix} 0 & -\sum_{j \in D_i} T_{ji} & 0 & 0 & 0 \\ 1/M_i & -D_i/M_i & 1/M_i & 0 & 0 \\ 0 & 0 & -1/T_{di} & 1/T_{di} & 0 \\ 0 & -1/(T_{gi}R_{gi}) & 0 & 1/T_{gi} & K_i/T_{gi} \\ 1 & -b_i & 0 & 0 & 0 \end{pmatrix}, B_{ji} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ \frac{T_{ji}}{M_i \sum_{h \in D_i} T_{ih}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$B_i = \begin{pmatrix} 0 & 1/M_i & 0 & 0 & 0 \end{pmatrix}^T,$$

where M_i and D_i are inertia and damping constant respectively, T_{gi} and T_{di} are the governor and gas turbine constant respectively, and R_{gi} and T_{ij} are the regulation and synchronizing constant respectively [20]. The related parameters of sub-system are listed in Table 1 [24].

Table 1. Parameters of sub-system.

Parameter Name	Description	Value
M_i	inertia constant	0.2
D_i	damping constant	0.26
T_{ji}	synchronizing constant	0.5
T_{di}	governor constant	5
R_{gi}	regulation constant	0.5
b_i	frequency bias gain	1
T_{gi}	gas turbine constant	0.2
K_i	tie-line bias control gain	0.1

The constraints of DoS attacks are set as follows:

$$\sum_{i=1}^{14} \sum_{j \in N_c(i)} \delta_{ji} \leq 1.4.$$

The subsystems in grid communicate with each other by communication channels. The above constraint means that the sum of all communication channel jamming possibilities is 1.4, or the average jamming possibility is 0.1 for each communication channel. The larger C in Equation (26) is, the more cyber resources for attackers would be taken. In the following simulations, we set the sampling period to 0.1 s.

5.1. Case I: IEEE 14 Bus Test System

The IEEE 14 bus test system is carried out under reinforcement learning control and DoS attacks. The initial state of power grid simulation is assumed to be affected by a large disturbance. The simulation time lasts 5 s. Without any control, we can see that the system collapses under serious DoS attacks after a large power grid disturbance occurs as shown in Figure 3. The frequency increases to a large value, which reaches to a maximal value about 40 Hz in 5 s. Therefore, a controller is required to maintain the stability. The parameters of reinforcement learning algorithm are listed in Table 2. The detail of parameter selection principle can be referred to [2].

Table 2. Parameter of the controller in the IEEE 14 bus test system.

Parameter Name	Description	Value
α	damping factor for cost p_i	0.9
N	Horizon length for cost	10
β_i	Learning rate of neural network	0.1
γ_i	Weight of group sparse regulation term	0.012
$ Li $	Norm of damping parameters L_i	0.1

The parameter β_i decides the result of IEEE 14 bus test system. If β_i is too large, the online learning of Equation (20) diverges; otherwise, too small β_i may cause the slow convergence rate. The selection of γ_i is another key factor. Too large γ_i may result in the poor control performance and high sparsity of W_i .

Figure 3. Frequency deviation of IEEE 14 bus system without control.

Figures 4 and 5 show the results that use the proposed control scheme and the optimal DoS attack strategy. The optimal DoS attack strategy is obtained by solving Equation (26), and the optimal controllers are learned by the online iteration of Equation (20). According to the results in Figure 4, we can see the frequency deviation curves of all sub-systems converge to a steady value around 0. The maximal fluctuant magnitude of these curves is a small value about 1.2 Hz. The swings end in about 4 s.

For the optimal DoS attack strategy shown in Figure 5a, the attacks focus on the communication from buses 3, 6, 10, 11, and 13. For the optimal control structure about placement of sensor, actuator, and communication topology shown in Figure 5b, it shows the F-norm for block W_{ij} in W_i , $i, j = 1, 2, \dots, n$. If the norm value is large, it means sub-system j needs information from sub-system i . Thus, sub-system i needs to install sensors, and sub-system j requires actuators as well as the communication topology from sub-system i to j . The sensors are mainly installed in buses 2, 3, 6, 10, 11, 13, and 14. According to the solution, those communication lines that are not attacked maintain the system stable. Thus, the obtained solution is optimal for both attacker and defender (control system designer) as Nash equilibrium.

Figure 4. Frequency deviation of IEEE 14 bus system under control.

Figure 5. Optimal DoS attacks and control structure 3D Graph in the IEEE 14 bus system.

5.2. Case II: IEEE 24 Bus Test System

This section verifies the effectiveness of the proposed scheme in a relatively large system—IEEE 24 bus test system. The simulation time lasts 10 s. Both reinforcement learning control and DoS attacks are applied. The optimal attack strategy and the optimized control structure are obtained by solving

the optimization of Equation (26). Specifically, the optimal attack strategy is solved by the offline optimization algorithm described in Section 4.2, and the optimized control structure is obtained by the online learning of Equation (20).

The parameters of reinforcement learning algorithm are listed in Table 3. The detail of parameter selection can be referred to [2].

Table 3. Parameter of controller in IEEE 24 bus test system.

Parameter Name	Description	Value
α	damping factor for cost p_i	0.8
N	Horizon length for cost	10
β_i	Learning rate of neural network	1
γ_i	Weight of group sparse regulation term	0.005
$ Li $	Norm of damping parameters L_i	0.1

The simulation results are shown in Figures 6 and 7. Figure 6 shows the frequency deviation curves of power grid under the worst case of DoS attacks. The results illustrate the effectiveness of propose solution under the large-size disturbance and DoS attacks. The swings of all the frequency deviation curves end in about 5 s. The magnitude of swings is small, and its maximal value is 1.5 Hz.

Figure 6. Cont.

Figure 6. Frequency deviation of IEEE 24 bus system under control.

From the optimal attack strategy results shown in Figure 7a, we know the attacker should focus on the communications from buses 2, 3, 4, 14, 19, 21, and 22 with the attack possibility of 0.2–0.35. The optimal control structure under the worst case of DoS attacks is shown in Figure 7b. The sensor should be installed in buses 2, 3, 4, 6, 14, 15, 19, 22, 23, and 24. The weight of bus 19 is high, which means the information obtained from bus 19 has high importance. Thus, DoS attacks have such an attack strategy that attacks the communication lines of bus 19 with a relatively high attack rate.

Figure 7. Optimal DoS attacks and control structure 3D graph in the IEEE 24 bus system.

6. Conclusions

This paper proposes a novel optimization method of control structure as well as a reinforcement learning method in an integrated neural network under DoS attacks. The reinforcement learning, which involves the frequency damping and control performance, is approximated by the integrated neural network. The frequency damping includes the desired control input holding the system stable and accelerating the convergence rate in online learning. The approximation of control performance is adopted as the reinforcement signal to enhance the long-term system performance. To obtain a reasonable control scheme and structure under DoS attacks, a Stackelberg game model is also proposed. The worst case of DoS attacks is considered and solved by the proposed optimization solution. The optimization solution also derives the optimal control structure, which involves the placement of sensors, actuators (DES), and the communication topology. The simulation results illustrate the

effectiveness of the proposed scheme. The optimal DoS attacks and the control structure are obtained in the simulation. In future, the convexity of the game and the existence of game equilibrium will be analyzed. The constraints of DER considering the environment and variation will also be considered. In addition, the error of the proposed algorithm will be analyzed theoretically and numerically in the next steps.

Author Contributions: Conceptualization, J.S.; methodology, J.S.; software, J.S.; validation, J.S., G.Q. and Z.Z.; formal analysis, Z.Z.; investigation, G.Q. and Z.Z.; resources, J.S.; data curation, J.S. and Z.Z.; writing—original draft preparation, J.S.; writing—review and editing, G.Q.; visualization, Z.Z.; supervision, G.Q.; project administration, J.S. and Z.Z.; funding acquisition, J.S. and Z.Z..

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 61803061 and 61703347); Fundamental Research Funds for the Central Universities (Grant No. XDJK2019C019); Science and Technology of the Chongqing Natural Science Foundation (Grant No. cstc2016jcyjA0428); the Research Program of Chongqing Municipal Education Commission (Grant No. KJQN201800603); the Innovation Project of Chongqing Overseas Students Entrepreneurial Innovation Support program (Grant No. cx2018074); Chongqing Key Industries Common Key Technology Innovation project (Grant No. cstc2017zdcy-zdyf0366); and Southwest University Education Reform Project (Grant No. 2017JY080).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Chen, P.; Li, J.; Fei, M.R. Resilient event-triggered H_{∞} load frequency control for networked power systems with energy-limited DoS attacks. *IEEE Trans. Power Syst.* **2017**, *32*, 4110–4118.
- Sun, J.; Zhu, Z.; Li, H.; Chai, Y.; Qi, G.; Wang, H.; Hu, Y.H. An integrated critic-actor neural network for reinforcement learning with application of DERs control in grid frequency regulation. *Int. J. Electr. Power Energy Syst.* **2019**, *111*, 286–299. [[CrossRef](#)]
- Sridhar, S.; Hahn, A.; Govindarasu, M. Cyber-Physical System Security for the Electric Power Grid. *Proc. IEEE* **2011**, *100*, 210–224. [[CrossRef](#)]
- Liang, G.; Zhao, J.; Luo, F.; Weller, S.; Dong, Z.Y. A Review of False Data Injection Attacks Against Modern Power Systems. *IEEE Trans. Smart Grid* **2017**, *8*, 1630–1638. [[CrossRef](#)]
- Yuan, Y.; Yuan, H.; Lei, G.; Yang, H.; Sun, S. Resilient Control of Networked Control System under DoS Attacks: A Unified Game Approach. *IEEE Trans. Ind. Inform.* **2016**, *12*, 1786–1794. [[CrossRef](#)]
- Srikantha, P.; Kundur, D. Denial of service attacks and mitigation for stability in cyber-enabled power grid. In Proceedings of the Innovative Smart Grid Technologies Conference, Washington, DC, USA, 18–20 February 2015.
- An, L.; Yang, G.H. Decentralized Adaptive Fuzzy Secure Control for Nonlinear Uncertain Interconnected Systems Against Intermittent DoS Attacks. *IEEE Trans. Cybern.* **2019**, *49*, 827–838. [[CrossRef](#)] [[PubMed](#)]
- Sun, H.; Peng, C.; Yang, T.; Zhang, H.; He, W. Resilient control of networked control systems with stochastic denial of service attacks. *Neurocomputing* **2017**, *270*, 170–177. [[CrossRef](#)]
- Hu, S.; Yue, D.; Xie, X.; Chen, X.; Yin, X. Resilient Event-Triggered Controller Synthesis of Networked Control Systems Under Periodic DoS Jamming Attacks. *IEEE Trans. Cybern.* **2018**. [[CrossRef](#)]
- Li, Y.; Quevedo, D.E.; Dey, S.; Ling, S. SINR-based DoS Attack on Remote State Estimation: A Game-theoretic Approach. *IEEE Trans. Control Netw. Syst.* **2017**, *4*, 632–642. [[CrossRef](#)]
- Yuan, Y.; Sun, F.; Liu, H. Resilient control of cyber-physical systems against intelligent attacker: A hierarchal stackelberg game approach. *Int. J. Syst. Sci.* **2016**, *47*, 2067–2077. [[CrossRef](#)]
- Ding, K.; Dey, S.; Quevedo, D.E.; Ling, S. Stochastic Game in Remote Estimation under DoS Attacks. *IEEE Control Syst. Lett.* **2017**, *1*, 146–151. [[CrossRef](#)]
- Ding, K.; Li, Y.; Quevedo, D.E.; Dey, S.; Ling, S. A multi-channel transmission schedule for remote state estimation under DoS attacks. *Automatica* **2017**, *78*, 194–201. [[CrossRef](#)]
- Zhu, Q.; Basar, T. Game-Theoretic Methods for Robustness, Security, and Resilience of Cyberphysical Control Systems: Games-in-Games Principle for Optimal Cross-Layer Resilient Control Systems. *IEEE Control Syst.* **2015**, *35*, 46–65.

15. Atwa, Y.M.; El-Saadany, E.F. Optimal Allocation of ESS in Distribution Systems With a High Penetration of Wind Energy. *IEEE Trans. Power Syst.* **2010**, *25*, 1815–1822. [[CrossRef](#)]
16. Borges, C.L.T.; Falcão, D.M. Optimal distributed generation allocation for reliability, losses, and voltage improvement. *Int. J. Electr. Power Energy Syst.* **2006**, *28*, 413–420. [[CrossRef](#)]
17. Zhang, H.; Ayoub, R.; Sundaram, S. Sensor selection for Kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms. *Automatica* **2017**, *78*, 202–210. [[CrossRef](#)]
18. Gupta, V.; Chung, T.H.; Hassibi, B.; Murray, R.M. On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica* **2006**, *42*, 251–260. [[CrossRef](#)]
19. Qu, C.; Chen, W.; Song, J.B.; Li, H. Distributed Data Traffic Scheduling With Awareness of Dynamics State in Cyber Physical Systems With Application in Smart Grid. *IEEE Trans. Smart Grid* **2015**, *6*, 2895–2905. [[CrossRef](#)]
20. Zhu, Z.; Sun, J.; Qi, G.; Chai, Y.; Chen, Y. Frequency Regulation of Power Systems with Self-Triggered Control under the Consideration of Communication Costs. *Appl. Sci.* **2017**, *7*, 688. [[CrossRef](#)]
21. Li, H. Data traffic scheduling for cyber physical systems with application in voltage control of microgrids. *IEEE Syst. J.* **2017**, *8*, 542–552. [[CrossRef](#)]
22. Cambiaso, E.; Papaleo, G.; Aiello, M. Slowcomm: Design, development and performance evaluation of a new slow DoS attack. *J. Inf. Secur. Appl.* **2017**, *35*, 23–31. [[CrossRef](#)]
23. Cambiaso, E.; Papaleo, G.; Giovanni, C.; Aiello, M. A Network Traffic Representation Model for Detecting Application Layer Attacks. *Int. J. Archit. Comput.* **2016**, *5*, 31–42.
24. Sun, J.; Li, J. A Stable Distributed Neural Controller for Physically Coupled Networked Discrete-Time System via Online Reinforcement Learning. *Complexity* **2018**, *2018*, 5950678. [[CrossRef](#)]
25. Ding, D.; Wang, Z.; Ho, D.W.; Wei, G. Observer-Based Event-Triggering Consensus Control for Multiagent Systems With Lossy Sensors and Cyber-Attacks. *IEEE Trans. Cybern.* **2017**, *47*, 1936–1947. [[CrossRef](#)] [[PubMed](#)]
26. Robnik-šikonja, M. Data Generators for Learning Systems Based on RBF Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 926–938. [[CrossRef](#)] [[PubMed](#)]
27. Hansen, L.P.; Sargent, T.J. Robust Control and Model Uncertainty. *Am. Econ. Rev.* **2001**, *91*, 60–66. [[CrossRef](#)]
28. Pan, C.; Liu, W.; Thompson, J.S.; Yang, C.; Jorswieck, E.A. Semi-dynamic Green Resource Management in Downlink Heterogeneous Networks by Group Sparse Power Control. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 1250–1266.
29. Bertsekas, D.P. *Constrained Optimization and Lagrange Multiplier Methods*; Academic Press: Cambridge, MA, USA, 1982.
30. Polak, E.; Yang, T.H.; Mayne, D.Q. A Method of Centers Based on Barrier Functions for Solving Optimal Control Problems with Continuum State and Control Constraints. *SIAM J. Control Optim.* **2006**, *31*, 159–179. [[CrossRef](#)]
31. Zimmerman, R.D.; Murillo-Sanchez, C.E.; Thomas, R.J. MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education. *IEEE Trans. Power Syst.* **2011**, *26*, 12–19. [[CrossRef](#)]