Complexity

Theory and Applications of Cyber-Physical Systems

Lead Guest Editor: Qiuye Sun
Guest Editors: Rui Wang, Renke Han, and Yushuai Li

Theory and Applications of Cyber-Physical Systems

Complexity

Theory and Applications of Cyber-Physical Systems

Lead Guest Editor: Qiuye Sun
Guest Editors: Rui Wang, Renke Han, and Yushuai Li

Contents

Hindawi
Complexity
Volume 2022, Article ID 5995281, 15 pages
https://doi.org/10.1155/2022/5995281

Research Article

Distributed Energy Management for Port Power System under False Data Injection Attacks

Qihe Shan,[1] **Xin Zhang**,[1] **Qiongyue Zhang**,[1] **and Qiuye Sun**[2]

[1]*Navigation College, Dalian Maritime University, Dalian 116026, China*
[2]*School of Information Science and Engineering, Northeastern University, Shenyang 110819, China*

Correspondence should be addressed to Qiuye Sun; sunqiuye@ise.neu.edu.cn

Received 30 April 2021; Revised 9 December 2021; Accepted 24 December 2021; Published 31 January 2022

Academic Editor: Ning Cai

This paper investigates a distributed energy management strategy for the port power system under false data injection attacks. The attacker can tamper with the interaction information of energy equipment, penetrate the boundary between port information system and port power system, and cause serious operation failure of port energy equipment. Firstly, a hierarchical topology is proposed to allocate the security resources of the port power system. Secondly, by reconstructing the topological structure of the port information system, the robustness of the information system is improved, and the impact of false data injection attacks on the port power system is reduced, which realizes secure distributed energy management of the port power system. Finally, the simulation results show the effectiveness of the proposed strategy, and the defense capability of the port power system is improved.

1. Introduction

With the development of the shipping industry [1], the types and quantities of energy equipment in port are increasing, and the port has become an energy-intensive region that relies on fossil energy; pollutant emissions are also increasing year by year [2]. In order to achieve effective pollution control, the interaction of multiple stakeholders in the maritime shore power system is required [3]. Port enterprises should speed up the realization of the strategy of "implementing shore power," and the government should actively supervise pollution emissions and encourage the use of clean energy [4]. Optimizing the utilization rate of renewable energy is the breakthrough point to solve the problem of high energy consumption and high emission in ports, but as the penetration rate of renewable energy becomes higher and higher, its randomness and volatility bring uncertainty to the port power system [5, 6]. More noteworthy, the Port of Barcelona was the target of a cyberattack that affected some of its servers and systems on the morning of September 20, 2018, forcing the organization to launch the contingency plan designed specifically for these incidents [7]. An Israeli cyber attack on the Iranian port setup caused serious chaos in the port's waterways and roads according to the Washington Post reported in May 2020 [8]. A study by Lloyd's of London showed that a cyberattack affecting major ports in the Asia-Pacific could cost $110 billion [9]. Therefore, how to carry out port energy management has become a key issue in order to make full use of the renewable energy of the port, reduce the environmental pollution caused by the redundant power supply, and ensure the normal run of the port power system.

At present, the energy management of port mainly draws on the experience and methods of the distribution grid, and it is actually an optimization problem of the complex system considering various practical constraints [10]. It aims to realize the balance of supply and demand and to maximize the operation benefit of the port power system when meeting the physical constraints of energy equipment [11, 12]. At the same time, the ports that implement wind power generation and photovoltaic power generation have the characteristics of distribution and uncertainty, which brings difficulties to the traditional centralized energy management strategy [13, 14]. Port energy management is gradually transformed into a distributed management method with the advantages of

flexibility, high reliability, and scalability [15–17]. A number of studies concerned about energy management have been proposed for power systems, including hierarchical decentralized energy management, multiagent system-based coordinated operation strategy, and hybrid energy management [18–20]. To solve the energy management problem of the energy Internet, the work in [18] proposed a completely distributed algorithm, in which each participant can realize the estimation of the optimal energy price and the calculation of the optimal energy output by only interacting with the neighbors. The large-scale connected inverter (CI) networks increase the reliability and resilience for the power grid, which may also cause grievous adventure to the reliable and stable operation of the main grid if the control of multiple CIs could not enable good cooperation [21, 22]. Lu et al. proposed a master-slave cluster cooperation strategy over the two-layer switching cyber network topologies to realize the economic distribution among multiple clusters [23]. There have been many studies on ports in different directions. The literature [24–26] reduces the peak load in the port by excavating the peak cutting and valley filling capacity of port freezers, terminal cranes, and electric vehicles. In reference [27], aiming at the problem that a large number of flexible loads such as reefers, ships, and electric vehicles are difficult to be highly coordinated, the port load management optimization is realized by using multiagent system with electricity price as the regulation signal. A novel decentralized power management method for a large port-based multiagent system (MAS) is proposed in [28].

It is specially mentioned that the information interaction process is vulnerable to malicious network attacks among the calculation and analysis of energy supply equipment due to the strong distributed characteristics of the current port power system [29]. In the next place, owing to the rapid development of information technology, there is a strong coupling characteristic between port cyber system and its power system [30, 31], which determines that the secure operation of information communication network is a basic foundation for the normal operation of its power system obviously [32]. Considering the increasing degree of dependence between information flow and power flow, the threat on information communication flow within the port power system will influence the normal operation of the power system, which will penetrate the boundary between the above two systems and have several negative impacts on disturbance, system instability, and collapse of the energy system [33, 34]. Thus, in order to ensure the reliable operation of the port power system under the above situations, it is necessary to strengthen the security defense of its information and communication system, so as to ensure the normal operation of its internal equipment [35–37]. It is noted that the operations security and efficiency of energy utilization are the most significant requirements within the port cyber-physical system [38]. Therefore, ports must make cyber security a top priority to ensure its security, compliance, and commercial competitiveness. The direct target of network attacks is the information flow in the information layer. It is common that cyberattacks include denial of service attacks, packet loss attacks, false data injection attacks (FDIAs), malicious software viruses, and so on [39]. Among them, the false data injection attack approach is easy to implement and has the characteristics of strong concealment and interference, which have attracted widespread attention from researchers in related fields. Literature [40] mainly treats the perturbation term in FDI attacks as a constant. In Literature [41], to verify the robustness of the proposed technique, four cases of FDI attacks across attacker injection signals were considered that are non-periodic attack, non-periodic replacement attack, periodic attack, and simultaneous attack. All of the above attack types will be within our consideration. In the next place, we need to focus on existing ways to resist attacks, a method for detecting FDIA in DC microgrid current measurement based on distributed control strategy control is introduced in [42]. The authors study fault identification using SVM, decision trees, and random forests in [43, 44]. The detection method of FDIA was studied in the above literature. The work in [45] proposes an antioffensive cooperative control strategy, which can adjust the power of the virtual power plant under a specific scheduling command. References [46, 47] proposed a trust-based antiattack resilient cooperative distributed control method and designed a resilient synchronization protocol to solve the problem of sensor attacks and reduce the adverse effects of attacks on communication links and hijacking of controllers. The author analyzed the relationship of the maximum amount of tolerable attacks and the number of total agents, and connected number has been provided for consensus under adversarial attacks, and a sequence of resilient consensus algorithms were developed in [48, 49].

However, the existing defense mechanisms rely mainly on the means of detecting attacks, and the effectiveness of those is still restricted by the maximum amount of tolerable attacks. To actively defend and relax the assumption on the maximum amount of tolerable attacks, this article was completed inspired by [50]. This paper proposes a distributed energy management strategy for the port power system based on hierarchical topology reconstruction of communication networks in order to reduce the impact of false data injection attacks on port energy management within limited defense resources to ensure the safe run of the system with lower security defense cost.

The main contributions of this paper are summarized as follows:

(1) This paper investigates the problem of distributed port energy management when the network of the port power system is under attack, where the tolerable number of attack nodes is unknown and can be arbitrarily large.

(2) To increase the security capacity of the port power system under false data injection attacks, this paper proposes a hierarchical topology reconstruction method of the port information system, which can reduce the impact of false data injection attacks on the port power system during the whole operation condition.

The rest of the paper is organized in the following. Section 2 presents the main types of port power facilities and energy management. The secure distributed energy management under FDIAs is detailed given in Section 3. In Section 4, the proposed method is simulated for port power system, and the obtained results are presented and evaluated. Finally, Section 5 draws a general conclusion.

2. Distributed Energy Management of Port Power System

2.1. Port Power System Configuration. In the port information energy system, there are multiple energy supply, energy demand, and energy storage entities [51]. In terms of energy supply, the port is particularly suitable for converting natural resources such as wind energy, tidal energy, and solar energy into electrical energy; in terms of energy use, the port uses a large number of cranes, gantry cranes, bridge cranes, plug-in electric vehicles, and so on [52]. In terms of energy storage, it is necessary to alleviate the uncertainty of port load and absorb intermittent and fluctuating renewable energy; energy storage equipment needs to be considered in the port power system.

2.1.1. Renewable Generator. Solar and wind energy are the main renewable energy sources. Considering the intermittency and volatility of renewable energy, it is unable to participate in the energy management process of the port, so according to the dispatch forecast curve before the day, the average value is used as a reference. Taking renewable power generation equipment as an example, its power generation can be expressed as follows:

$$\widetilde{P}_i = \frac{\int_t^{t+T} P_i dt}{T}, \tag{1}$$

where $\widetilde{p}_i$ is the power generation predicted by the renewable energy equipment i according to the day-ahead dispatch period T. Assuming that the prediction error described by the probability density function obeys the Gaussian distribution, it can be expressed as follows:

$$f_i(\Delta P_i) = \frac{1}{\sqrt{2\pi}\delta_i} e^{-\left(\Delta P_i/2\delta_i\right)^2}. \tag{2}$$

Choose an appropriate confidence level to get the confidence interval of the prediction error $[\Delta P_i^{\min}, \Delta P_i^{\max}]$. Thus, the power generation of renewable energy equipment i can be expressed as follows:

$$\begin{aligned} &\underline{P}_i \le P_i \le \overline{P}_i, \\ &P_i = \widetilde{P}_i + \Delta P_i. \end{aligned} \tag{3}$$

To sum up, considering the actual operation effect, the operation cost function of renewable energy equipment i can be expressed as follows:

$$C_i(P_i) = a_i {P_i}^2 + b_i P_i + c_i, \tag{4}$$

where b_i and c_i are the coefficients of the operating cost function; the optimality and possibility of renewable energy equipment operation can be balanced to take renewable energy into energy management.

2.1.2. Fuel Generator. The use of alternative clean energy in port equipment and buildings can effectively reduce environmental pollution and greenhouse gas emissions in the port area, which is in line with the development concept of green port [48]. The choice of alternative clean energy fuels plays an important role in the sustainable development of ports and the improvement of the port environment. The local optimization problem of fuel energy equipment can be transformed into

$$\min C_i(P_i) = a_i(P_i)^2 + b_i P_i + c_i + \tau_i \exp(\eta_i P_i), \tag{5}$$

$$P_i^{\min} \le P_i \le P_i^{\max}, \tag{6}$$

$$-P_i^{\text{ramp}} \le P_i(k) - P_i(k-1) \le P_i^{\text{ramp}}. \tag{7}$$

2.1.3. Shore-Ship Power Supply. In order to effectively promote the development of energy-saving and low-carbon emission reduction, the use of shore-based power supply by ships calling at ports is one of the key tasks of the port industry for energy conservation and emission reduction. When the ship is at the berth, the ship's auxiliary generator runs in parallel with the shore power source to ensure the ship's reliable operation. The operating cost of the auxiliary engine (only fuel is considered) is approximated by the second-order polynomial of the power generated by the generator [28]:

$$U_i(P_i) = a_i(P_i)^2 + b_i P_i + c_i. \tag{8}$$

2.1.4. Plugged in Electric Vehicle. Electronic transportation in the port improves energy efficiency and reduces greenhouse gas emissions. The use of electric vehicles instead of port internal combustion engine operating vehicles (such as automobiles, forklifts, etc.) will effectively reduce carbon emissions in the port. In the future, large ports can charge and discharge electric vehicles through charging piles, thereby increasing the flexibility of providing port energy requirements. The cost function of PEV can be modeled as follows [27]:

$$\max U_i(P_i) = -a_i(P_i + b_i)^2. \tag{9}$$

Electric vehicles are constrained by battery charging efficiency and battery capacity as follows:

$$E_r(k) - E_r(k-1) = \begin{cases} \varsigma^{ch} P_i T \\ \\ \dfrac{1}{\varsigma^{\text{ds}}} P_i T \end{cases}. \tag{10}$$

2.1.5. Distributed Power Storage Device. Distributed renewable energy generation has the characteristics of intermittent and volatility. Large-scale access to distributed power generation equipment increases the adjustment difficulty of the grid and is likely to cause a lot of waste of resources such as abandoning light and wind; When a large number of short-cycle loads are connected, it will cause large fluctuations in the frequency or voltage of the grid, which will affect the stability of other electrical equipment. The addition of energy storage system in the power grid will not only help to reduce the abandonment of new energy and make up for the instability of new energy power generation but also reduce the demand pressure at the peak of power consumption by releasing the stored energy. On the energy-consuming side, it meets the rapid response requirements of frequency and voltage regulation and provides high-quality electric energy. The cost of energy storage equipment i can be defined as follows:

$$C_i(P_i) = a_i(P_i + b_i)^2, \tag{11}$$

where p_i represents the charge and discharge power of the energy storage device i, and its symbol represents the state of charge and discharge (the discharge state is positive, and the charge state is negative). Energy storage equipment cannot work in charging and discharging at the same time and needs to meet multiple restricted operating conditions:

$$-P_i^{ch,\max} \le P_i \le P_i^{\mathrm{ds},\max}, \tag{12}$$

$$SoC(k) - SoC(k-1) = \begin{cases} \varsigma^{ch} P_i T, & p_i \le 0 \\ \dfrac{1}{\varsigma^{\mathrm{ds}}} P_i T, & p_i > 0 \end{cases}, \tag{13}$$

where P_i should be between the maximum charging power $P_i^{ch,\max}$ and the maximum discharging power $P_i^{\mathrm{ds},\max}$, ς^{ch} and ς^{ds} are positive numbers between 0 and 1, and represents the energy loss during the charging and discharging process of the energy storage device i.

3. Distributed Energy Management Scheme of Port Power System

The goal of energy management of the port power system is to maximize the operation benefit or minimize the operation cost on the basis of ensuring the safe operation of port power equipment and the balance of supply and demand of the port power system.

We assume that the port power system can be represented by an undirected graph $G = (V, E, A)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is a vertex set, $E \subseteq V \times V$ is the set of undirected edges, and A is weighted and non-negative adjacency matrix [50]. It assumed that there are n power devices in the port power system, whose cost function of device i can be denoted by $C_i(P_i)$, where P_i is the output power of device i. The energy management problem can be expressed as follows:

$$\max \sum_{i=1}^{n} -C_i(P_i), \tag{14}$$

subject to the following constraints:

(a) Device constrains

$$p_i^{\min} \le p_i \le p_i^{\max}, \tag{15}$$

$$P_4^* = 40.0000\,\mathrm{MW}, \tag{16}$$

where $P_i^{\min}, P_i^{\max}$ are the minimum and maximum power output of the device i, respectively, and P_i^{ramp} is the ramp rate constraint of the power facility.

(b) Supply-demand balance constraints

$$\sum_{i=1}^{n} P_i = \sum_{i=1}^{n} d_i, \tag{17}$$

where d_i is load. It is assumed that the energy management problem is solvable; the following conditions are satisfied:

$$\sum_{i=1}^{n} P_i^{\min} \le \sum_{i=1}^{n} d_i \le \sum_{i=1}^{n} P_i^{\max}. \tag{18}$$

Assumption 1. For each $i \in \{1, 2, \ldots, N\}$, the cost function $C_i(P_i)$: $R_+ \longrightarrow R_+$ is strictly convex and continuously differentiable, where R_+ denotes the set of non-negative real numbers.

The Lagrangian function of problem (14) and constraints (15) is as follows:

$$\begin{aligned} L(P, \lambda) &= \sum_{i=1}^{n} \left(-C_i(P_i)\right) - \lambda \left(\sum_{i=1}^{n} d_i - \sum_{i=1}^{n} P_i \right) \\ &= \sum_{i=1}^{n} \left(-C_i(P_i) + \lambda P_i\right) - \lambda \sum_{i=1}^{n} d_i, \end{aligned} \tag{19}$$

where λ is the dual variable. By decoupling P_i from $L(P, \lambda)$, we get

$$P_i^* = \arg\max \left(-C_i(P_i) + \lambda P_i\right). \tag{20}$$

Assumption 2. The Lagrangian function $L(P, \lambda)$ has a saddle point, that is, there exists an optimal solution (P^*, λ^*) such that

$$L(P^*, \lambda) \le L(P^*, \lambda) \le L(P^*, \lambda). \tag{21}$$

holds for all $P_i \in \Omega$, $\lambda \in R_+$.

With Assumption 2, problem (14) is a convex optimization problem, and the duality gap is 0; it can be transformed into its dual problem, which can be expressed as follows:

$$\min D(\lambda) = \min L(P^*, \lambda), \\ = \min\left(\sum_{i=1}^{n}(-C_i(P_i^*) + \lambda P_i^*) - \lambda\sum_{i=1}^{n} d_i\right). \quad (22)$$

Using the gradient descent method to solve, the update of $\lambda(k)$ can be described as follows:

$$\lambda(k+1) = \lambda(k) - \alpha\frac{\partial D}{\partial \lambda}(\lambda(k)), \quad (23)$$

where $\alpha > 0$ is a constant that represents the step size. The energy management based on the gradient descent method can be expressed as follows:

$$P_i(k) = \max\left\{P_i^{\min}, \min\left\{\arg\max\left(-C_i(P_i) + \lambda(k)P_i\right), P_i^{\max}\right\}\right\}, \quad (24)$$

$$\lambda(k+1) = \lambda(k) - \alpha\left(\sum_{i=1}^{n} P_i(k) - \sum_{i=1}^{n} d_i\right). \quad (25)$$

It is worth noting that the above energy management method is centralized and $\lambda(k)$ is a global variable, which requires the centralized control center to collect the information of all devices in the system and greatly increases the computing burden and communication burden of the control center. In this case, once the control center fails, the energy management process of the port power system will collapse, and the safe operation of the system will be seriously threatened. In [54], a distributed estimator $\lambda_i(k)$ is introduced to evaluate global variables $\lambda(k)$, and a distributed energy management algorithm is designed through a finite number of consensus protocol calculations, which alleviates the communication burden and computing burden of the scheduling center (Algorithm 1).

Here, $w_{ij} > 0$ is the weight assigned by node i to node j.

Assumption 3. The existence of a double random matrix $W = [w_{ij}]_{n\times n}$ which satisfies $W = W^T$ and for $(v_i, v_i) \notin E \cup (v_i, v_i)$, $w_{ij} \neq 0$.

Lemma 1. *Under the above assumptions, the iteration number of consensus updates meets* $\sigma \geq (\log\beta - \log(4n(\beta + \alpha l))/\log\gamma$, *and the initial value* $\lambda_i(0)$ *satisfies* $|\lambda_i(0) - \overline{\lambda}(0)| \leq \beta$; *sequence* $\{\lambda_i(k)\}$ *can be achieved consensus; and* $\{P(k)\}$ *can arbitrarily approach the optimal solution* P^*.

Remark 1. The above algorithm does not need centralized scheduling and decentralize the computation burden of the scheduling center. Each node only needs to exchange information with its neighbors to alleviate the communication burden of the central nodes. In this case, distributed scheduling method assumes that the network environment is benign, and if there are intruders in the information network, it is very likely that the consensus between nodes cannot be achieved.

4. Distributed Energy Management under FDIAs

With the deep fusion of information flow and energy flow in the port power system, the safe operation of the port power system is facing many challenges. More and more energy devices with communication abilities are connected to the port power system, and the information transmitted by energy equipment is threatened by network attacks [55].

4.1. Energy Management Problem under FDIAs. Considering the features of concealment and interference, false data injection attacks (FDIAs) have been regarded as one of the most popular attacks of the communication system in the port power system [56]. Furthermore, the attacker can track the configuration of the port power system, tamper with the sensing data of the sensor or tamper with the command signal sent by the controller, which will lead to the error of the state estimation process, and finally make the power system unable to operate normally [57]. As mentioned above, the false data injection attack model is as follows:

$$\lambda_i(k+1) = \zeta_i(\lambda_i(k)), \quad (26)$$

where $\lambda_i(k)$ denotes the information of device i exchanged with other devices and $\zeta_i(\lambda_i(k))$ represents arbitrary update function.

Without losing generality, we assume that node $i = 1, 2, \ldots, n_0$ is not attacked by false data injection attacks and node $i = n_0 + 1, n_0 + 2, \ldots n$ is attacked by false data injection attacks. The interaction of $\lambda_i(k)$ in information layer can be rewritten as follows:

Initialization: Each node i initializes $\lambda_i(0)$, step size α and calculation times σ
Iteration: $(k \geq 0)$
(1) Each node i updates $P_i(k)$ according to (24)
(2) Each node i updates $v_i^0(k+1)$ based on the gradient descent method, $v_i^0(k+1) = \lambda_i(k) - \alpha(P_i(k) - d_i)$
(3) Each node i interacts with its neighbors, until reached consensus. $v_i^1(k+1) = \sum_{v_j^0 \in S_i} w_{ij} v_j^0(k)$, $v_i^2(k) = \sum_{v_j^1 \in S_i} w_{ij} v_j^1(k)$, $v_i^\sigma(k) = \sum_{v_i^{\sigma-1} \in S_i} w_{ij} v_i^{\sigma-1}(k)$
(4) Each node i estimates $\lambda_i(k+1)$. $\lambda_i(k+1) = v_i^\sigma(k)$
Let $k = k+1$, turn to 1.

Algorithm 1: Distributed energy management scheme.

$$\begin{bmatrix} \lambda_1(k+1) \\ \lambda_2(k+1) \\ \vdots \\ \lambda_{n_0}(k+1) \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n_0} & w_{1(n_0+1)} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n_0} & w_{2(n_0+1)} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ w_{n_0 1} & w_{n_0 2} & \cdots & w_{n_0 n_0} & w_{n_0(n_0+1)} & \cdots & w_{n_0 n} \end{bmatrix} \begin{bmatrix} v_1^{\sigma-1}(k) \\ v_2^{\sigma-1}(k) \\ \vdots \\ v_{n_0}^{\sigma-1}(k) \\ v_{n_0+1}^{\sigma-1}(k) \\ \vdots \\ v_n^{\sigma-1}(k) \end{bmatrix}, \tag{27a}$$

$$\begin{bmatrix} \lambda_{n_0+1}(k+1) \\ \vdots \\ \lambda_n(k+1) \end{bmatrix} = \begin{bmatrix} \zeta_{n_0+1}\left(v_{n_0+1}^{\sigma-1}(k)\right) \\ \vdots \\ \zeta_n\left(v_n^{\sigma-1}(k)\right) \end{bmatrix}. \tag{27b}$$

The goal of energy management of the port power system under FDIAs is to maximize the operation benefit of non-attacked devices. The energy management problem can be expressed as follows:

$$\max \sum_{i=1}^{n_0} -C_i(P_i), \tag{28}$$

subject to the following constraints:

(a) Device constrains

$$p_i^{\min} \leq p_i \leq p_i^{\max}, \tag{29}$$

$$-P_i^{\text{ramp}} \leq P_i(k) - P_i(k-1) \leq P_i^{\text{ramp}}, \tag{30}$$

where $P_i^{\min}, P_i^{\max}$ are the minimum and maximum power output of the device i, respectively, and P_i^{ramp} is the ramp rate constraint of the power facility.

(b) Supply-demand imbalance constraints

$$\left| \sum_{i=1}^{n_0} P_i - \sum_{i=1}^{n_0} d_i \right| \leq P_{\text{thres}}^{\text{im}}, \tag{31}$$

where $P_{\text{thres}}^{\text{im}}$ is the constant threshold of power mismatch under false data injection attacks.

4.2. Secure Distributed Energy Management Based on Topology Reconfiguration. The topology of the information

network is the key factor to realize the coordinated operation of port power devices and energy management of port power systems [53]. It is noted that the reliable transmission of information is a necessary condition to ensure the safe, stable, and economic operation of the port energy system. Information networks can provide different services and guarantees for energy equipment in different regions. In order to build a safe information network environment, we consider changing the port information network from a flat topology to a hierarchical topology. The hierarchical topology is as Figure 1.

Hierarchical topology divides n nodes into m layers and marks the layers as 1 to m from top to bottom. The nodes in the first layer only send information to the second layer, and the nodes in the i layer at least send messages to one node in the $i+1$ layer. In each layer, the information interaction between nodes can be bidirectional. Using hierarchical topology, it is realized that when false data injection attacks occur in the port information network, we should divide the attacked node and the non-attacked node into different topology layers and reduce the weight of the non-attacked node assigned to the attacked node, so as to reduce the impact of network attacks on the port information network.

Inspired by the literature [50], we protect some nodes in the port power system from being invaded by attack nodes. By expanding the influence of these protected nodes in the network, to suppress the impact of false data injection attacks on the port information and energy system, the protection measures include: (1) improving the firewall security level of the protected nodes, and make redundant resources of the protected nodes to ensure that the protected nodes get real and reliable information; (2) Using the digital signature and data encryption technology, to the protected nodes the real information sent cannot be tampered. Before introducing a secure distributed energy management strategy, we introduce the following definitions and assumptions.

For the convenience of description, we regard the vertex set $T = \{v_i | i = 1, 2, \ldots n_1\}$ as the set of protected nodes, and the vertex set $A = \{v_i | i = n_0 + 1, \quad n_0 + 2, \ldots, n\}$ as the set of attacked nodes. The other vertex is ordinary nodes.

Remark 2. Topology reconstruction from communication network due to attack does not mean that the operation of the attacked node stops completely. The isolated node does not mean that its corresponding unit stops power output. It will continue to supply local load demand because the attack action of an isolated node only occurs at the communication layer. The physical layer topology of the actual power supply unit will not change due to the occurrence of a network attack.

The Lagrangian function of problem (28) and constraints (31) is as follows:

$$\begin{aligned} L(P, \widetilde{\lambda}) &= \sum_{i=1}^{n_0} \left(-C_i(P_i)\right) - \widetilde{\lambda}\left(\sum_{i=1}^{n_0} d_i - \sum_{i=1}^{n_0} P_i - P_{\text{thres}}^{\text{im}}\right), \\ &= \sum_{i=1}^{n_0} \left(-C_i(P_i) + \widetilde{\lambda} P_i\right) - \widetilde{\lambda}\left(\sum_{i=1}^{n_0} d_i - P_{\text{thres}}^{\text{im}}\right), \end{aligned} \tag{32}$$

where $\widetilde{\lambda}$ is the dual variable. By decoupling P_i from $L(P, \widetilde{\lambda})$, we get:

$$P_i^* = \arg\max\left(-C_i(P_i) + \widetilde{\lambda} P_i\right). \tag{33}$$

Remark 3. The above Lagrange function assumes that the supply is in short demand in the port power system. It can also assume that the supply exceeds the demand in the port power system. Owing to the port power system containing energy storage equipment, it can provide or store electric energy in time to reduce the negative pressure of power imbalance in case of emergency.

Assumption 4. The Lagrangian function $L(P, \widetilde{\lambda})$ has a saddle point, that is, there exists an optimal solution $(\widetilde{P}^*, \widetilde{\lambda}^*)$ such that

$$L\left(P, \widetilde{\lambda}^*\right) \le L\left(\widetilde{P}^*, \widetilde{\lambda}^*\right) \le L\left(P^*, \widetilde{\lambda}\right), \tag{34}$$

holds for all $P_i \in \Omega$, $\widetilde{\lambda} \in R_+$.

With Assumption 4, problem (14) is a convex optimization problem, and the duality gap is 0; it can be transformed into its dual problem, which can be expressed as follows:

$$\begin{aligned} \min D(\widetilde{\lambda}) &= \min L\left(P^*, \widetilde{\lambda}\right) \\ &= \min\left(\sum_{i=1}^{n_0}\left(-C_i(P_i^*) + \widetilde{\lambda} P_i^*\right) - \widetilde{\lambda}\left(\sum_{i=1}^{n_0} d_i - P_{\text{thres}}^{\text{im}}\right)\right) \\ &= \min \sum_{i=1}^{n_0} g_i(\widetilde{\lambda}). \end{aligned} \tag{35}$$

Before introducing a secure distributed energy management strategy, we introduce the following definitions and assumptions.

Definition 1. G_d is a connected dominating set of $G = (V, E, A)$ if each node i, which does not belong to the subset G_d, has at least one neighbor in the subset G_d and all nodes in G_d can form a connected graph.

Assumption 5. The protected nodes induce a connected dominating set of $G = (V, E, A)$.

Thus, a secure distributed energy management strategy based on topology reconfiguration is as follows (Algorithm 2,):

$\{\alpha(0), \ldots, \alpha(\infty)\}$ is the sequence of step sizes, which meets $\sum_{k=0}^{\infty} \alpha(k) = \infty$, $\sum_{k=0}^{\infty} \alpha^2(k) < \infty$, and $\alpha(k+1) \le \alpha(k)$.

The convergence of secure distributed energy management schemes under FDIAs topology reconstruction algorithm is given below.

According to the secure distributed energy management scheme under FDIAs, each node i determines the maximum threshold $\chi_i^M(k)$ and minimum threshold $\chi_i^m(k)$. And then, the set $R_i(k)$ can be obtained by threshold filtering. During the k iteration of node i, $R_i(k)$ can be divided into three

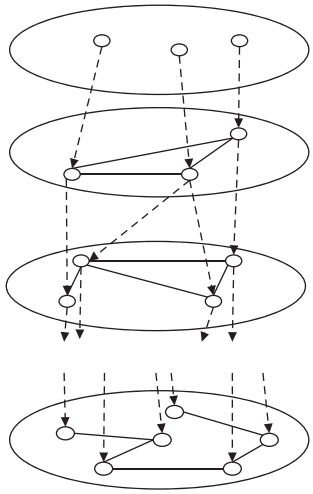


Figure 1: Hierarchical topology of information network.

Initialization: Each node i initializes $\lambda_i(0)$
Iteration: $(k \geq 0)$
(1) Each node i updates $P_i(k)$ according to (24)
(2) Each node i updates $\chi_i(k)$ based on the gradient descent method, $\chi_i(k) = \widetilde{\lambda}_i(k) - \alpha(k)(P_i(k) - d_i)$
(3) Each node i receives information, formulates the set $S_i(k) = \{\chi_i(k) | i \in N_i\}$
(4) Node i identifies the information of the protected node, compares with its own information, and then determines the maximum threshold $\chi_i^M(k)$ and minimum threshold $\chi_i^m(k)$
(5) Node i filters information between $\chi_i^m(k)$ and $\chi_i^M(k)$ and then formulates the set $R_i(k) = \left\{\chi_j(k) | \chi_i^m(k) \leq \chi_j(k) \leq \chi_i^M(k), \quad j \in N_i\right\}$
(6) Each node i updates $\chi_i(k+1)$, $m_{ij}(k) = 1/|R_i(k)|, \chi_j(k) \in R_i(k), \chi_i(k+1) = \sum_{\chi_j \in R_i(k)} m_{ij}(k)\chi_j(k)$
(7) Each node i estimates $\widetilde{\lambda}_i(k+1)$ and $\widetilde{\lambda}_i(k+1) = \chi_i(k+1)$
Let $k = k+1$, until $|\lambda_i(k) - \lambda_j(k)| < \varepsilon, \ j = 1, 2, \ldots, n_1$

Algorithm 2: Secure distributed energy management scheme under FDIAs.

types: from the set of protected nodes $R_i^T(k)$, itself $\chi_i(k)$, and the set of the other nodes $R_i^{\overline{T}}(k)$. Equation in Algorithm 2, can be rewritten as follows

$$\begin{aligned} \widetilde{\lambda}_i(k+1) &= \frac{1}{\left|R_i(k)\right|} \sum_{\chi_j(k) \in R_i(k)} \chi_j(k) \\ &= \frac{1}{\left|R_i(k)\right|} \left(\sum_{\chi_j(k) \in R_i^T(k)} \chi_j(k) + \chi_i(k) + \sum_{\chi_j(k) \in R_i^{\overline{T}}(k)} \chi_j(k) \right), \end{aligned} \tag{36}$$

where $\forall \chi_j(k) \in R_i^{\overline{T}}(k)$; it can be expressed as $\chi_j(k) = \rho_j \chi_i^m(k) + (1 - \rho_j)\chi_i^M(k), \quad 0 < \rho_j < 1$. Thus, equation in Algorithm 2, can be further rewritten as follows:

$$\begin{aligned}\widetilde{\lambda}_i(k+1) &= \frac{1}{\left|R_i(k)\right|}\left(\sum_{\chi_j(k)\in R_i^T(k)}\chi_j(k)+\chi_i(k)+\sum_{\chi_j(k)\in R_i^{\overline{T}}(k)}\chi_j(k)\right)\\ &= \frac{1}{\left|R_i(k)\right|}\left(\sum_{\chi_j(k)\in R_i^T(k)}\chi_j(k)+\sum_{\chi_j(k)\in R_i^{\overline{T}}(k)}\left(\rho_j\chi_i^m(k)+\left(1-\rho_j\right)\chi_i^M(k)\right)\right)\\ &= \sum_{\chi_j(k)\in R_i^T(k)\cup\chi_i(k)} m_{ij}\chi_j(k).\end{aligned} \tag{37}$$

Then, we have

$$\begin{bmatrix}\widetilde{\lambda}_1(k+1)\\ \vdots\\ \widetilde{\lambda}_{n_1}(k+1)\\ \vdots\\ \widetilde{\lambda}_{n_0}(k+1)\end{bmatrix}=\begin{bmatrix}m_{11} & \cdots & m_{1n_1} & \cdots & 0\\ \vdots & \ddots & \vdots & \cdots & \vdots\\ m_{n_11} & \cdots & m_{n_1n_1} & \cdots & 0\\ \vdots & \cdots & \vdots & \ddots & \vdots\\ m_{n_01} & \cdots & m_{n_0n_1} & \cdots & m_{n_0n_0}\end{bmatrix}\begin{bmatrix}\chi_1(k)\\ \vdots\\ \chi_{n_1}(k)\\ \vdots\\ \chi_{n_0}(k)\end{bmatrix}. \tag{38}$$

Compared with (27a), after topology reconstruction, the links from ordinary nodes to protected nodes and the links from attacked nodes to protected nodes are weakened, and the connection between attacking node and common node is also weakened. Equation (38) can be expressed as:

$$M(k)=\begin{bmatrix}M_{n_1\times n_1} & 0\\ M_{n_2\times n_1} & M_{n_2\times n_2}\end{bmatrix}_{n_0\times n_0}, \tag{39}$$

where $M_{n_1\times n_1}$ is the interaction between protected nodes, $M_{n_2\times n_1}$ denotes the interaction between protected node and ordinary nodes, and $M_{n_2\times n_2}$ is the interaction of itself. The nodes of the communication network can be divided into three layers by hierarchical topology reconstruction algorithm: layer 1 of trusted nodes, layer 2 of common nodes, and layer 3 containing attack nodes.

Furthermore, equation (38) can be written as follows:

$$\begin{aligned}\widetilde{\lambda}(k+1) &= M(k)\chi_i(k),\\ &= M(k)\left(\widetilde{\lambda}(k)-\alpha(k)\frac{\partial g}{\partial\widetilde{\lambda}}(\widetilde{\lambda}(k))\right),\\ &= M(k)\cdots M(0)\widetilde{\lambda}(0)-\sum_{t=0}^{k}M(k)\\ &\quad\cdots M(t+1)\alpha(t)\frac{\partial g}{\partial\widetilde{\lambda}}(\widetilde{\lambda}(t)),\\ &= \Psi(k,0)\widetilde{\lambda}(0)-\sum_{t=1}^{k+1}\Psi(k,t)\alpha(t-1)\frac{\partial g}{\partial\widetilde{\lambda}}(\widetilde{\lambda}(t-1)).\end{aligned} \tag{40}$$

Under Assumption 5 and Algorithm 2, $\Psi(k,t)$ have the following property: for any fixed t, n_2 non-zero entries in ψ(t) are lower bounded by φ^d.

To estimate $\widetilde{\lambda}(k)$, define the set as follows:

$$C(\mu,\eta)=\left\{h(\lambda)|h(\lambda)=\sum_{i=1}^{n_1}\gamma_i g_i(\lambda),\sum_{i=1}^{n_1}\gamma_i=1,\sum_{i=1}^{n_1}I\{\gamma_i\geq\mu\}=\eta\right\}, \tag{41}$$

and define the solution set as follows:

$$Y(\mu,\eta)=\cup_{h(\lambda)\in C(\mu,\eta)}\arg\min_{\lambda}h(\lambda). \tag{42}$$

By choosing $\mu=\varphi^d$ and $\eta=n_1$, it is easy to obtain $Y(\mu,\eta)$ is a convex set. According to reference [38], when $\alpha(k)$ meets $\lim_{k\longrightarrow\infty}\alpha(k)=0$, the $\lambda_i(k)$ of all protected nodes and ordinary nodes is convergent from (40), and converges to the set $Y(\mu,\eta)$ of optimal solutions of a weighted average of local cost functions $g_i(\widetilde{\lambda})$ belonging to all protected nodes and ordinary nodes.

Remark 4. Compared with literature [50], this paper focuses on secure energy management under FDIAs, which is a constrained optimization problem with coupling characteristics. By decoupling P_j from $L(P,\widetilde{\lambda})$, the dual problem is simplified, which make $\widetilde{\lambda}$ can achieve consensus.

5. Numerical Results

Consider port power system shown in Figure 2 and network topology under FDIAs shown in Figure 3; there are eight nodes in the port information network; the set of attacked nodes is $\{v_6,v_7,v_8\}$. The FDIA models are formed as

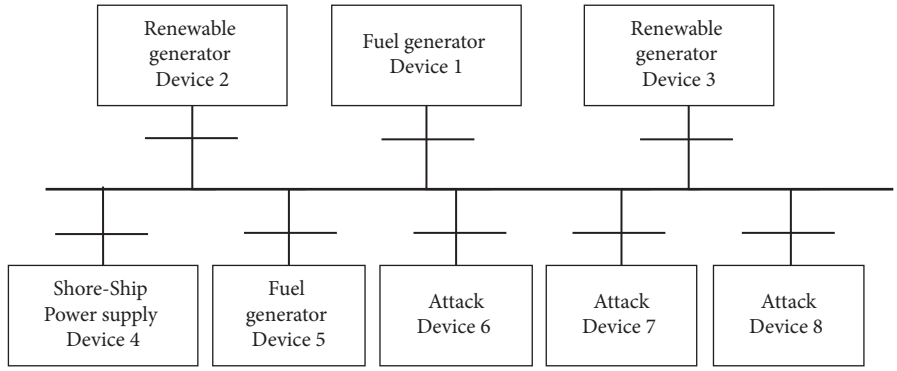


Figure 2: Port power system configuration.

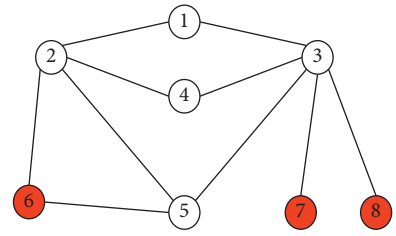


Figure 3: Network topology under FDIAs.

Table 1: The parameters of port power devices.

	Device 1	Device 2	Device 3	Device 4	Device 5
$a_i\,(\$/MW^2)$	0.040	0.032	0.023	0.054	0.040
$b_i\,(\$/MW)$	25	32	28	27	25
$c_i\,(\$)$	99	150	110	50	99

$\lambda_6(k+1) = 8\sin(0 \cdot 0025\pi\lambda_2(k+1)) + 35$, $\lambda_7(k+1) = (k/100)^2$, and $\lambda_8(k+1) = 40$.

We set cost function $C_i(P_i) = a_i P_i^2 + b_i P_i + c_i$ of devices i, with different parameters shown in Table 1. The loads are set as 85 MW, 50 MW, 105 MW, 85 MW, 77 MW, 30 MW, 33 MW, and 40 MW. According to Algorithm 1, set the initial value of the dual variable be 0, and the number of calculations is 40.

Figure 4(a) shows that when the nodes in the information network are attacked, the existing distributed energy management cannot guarantee the asymptotic consistency of the dual variable, and it leads to the outputs of power equipment fluctuating significantly, and the safe operation of the equipment is not realized as shown in Figure 4(b). As a result, the power mismatch of distributed energy management strategy cannot converge, and the imbalance constraint between supply and demand cannot be satisfied shown in Figure 4(c). It can be seen that the distributed energy management strategy loses effectiveness under FDIAs.

There are a total of eight nodes in the network of the information energy system: $\{v_1, v_2, v_3\}$ are protected nodes, $\{v_4, v_5\}$ are ordinary nodes, and $\{v_6, v_7, v_8\}$ are the nodes that are attacked. The network after topology reconstruction is shown in Figure 5, which divided nodes into two layers: the protected nodes are in the first layer and the other nodes are in the second layer. The information transfer from the first layer to the second layer in the topology is directed.

In the case of a network attack, the dual variables of nodes 1–5 are asymptotically consistent by topology reconstruction shown in Figure 6(a). Furthermore, Figure 6(b) shows that the output of all energy equipment tends to be safe run with small fluctuations, indicating that the energy strategy after topology reconstruction basically guarantees the safe run of the system. The energy management solution is given by $P_1^* = 73.8660$ MW, $P_2^* = 40.0000$ MW, $P_3^* = 62.9363$ MW, $P_4^* = 40.0000$ MW, and $P_5^* = 73.8660$ MW.

Power mismatches for secure distributed energy management fluctuate around −20 as shown in Figure 6(c), which greatly reduces the degree of energy mismatch compared with before the topology reconstruction. Although the energy mismatch still fluctuates to a small extent due to the direct transmission of information between nodes of the layered network after topology reconstruction, it is acceptable because it is safer than before topology reconstruction. In summary, the distributed optimization scheduling strategy has been topologically restructured to increase the robustness of the network and strengthen the active defense capability of the network.

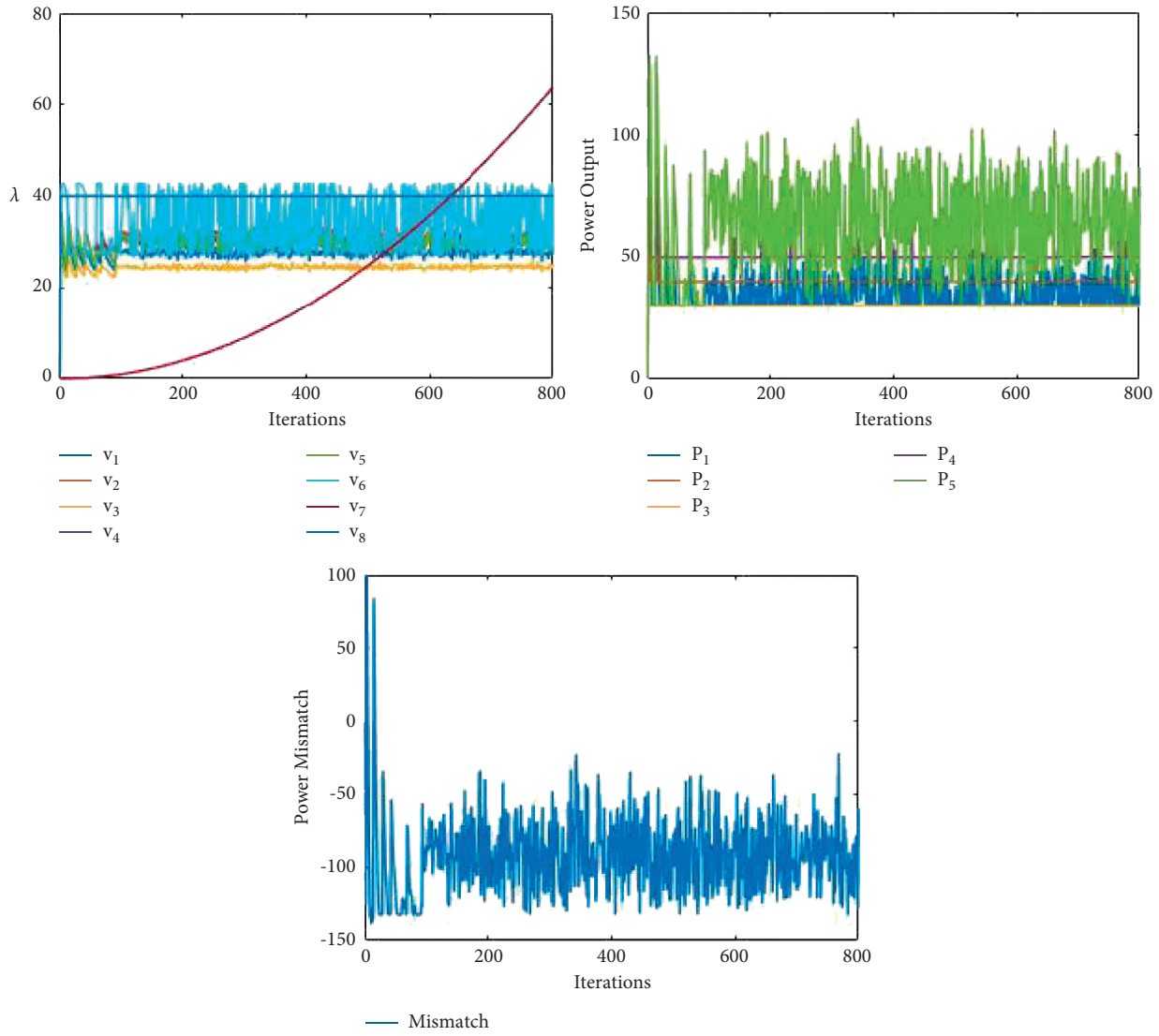


Figure 4: Variables of distributed energy management under FDIAs: (a) dual variables, (b) power output, and (c) power mismatch.

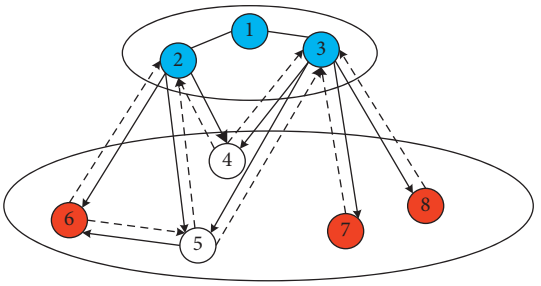


Figure 5: Reconfigurable network topology under FDIAs.

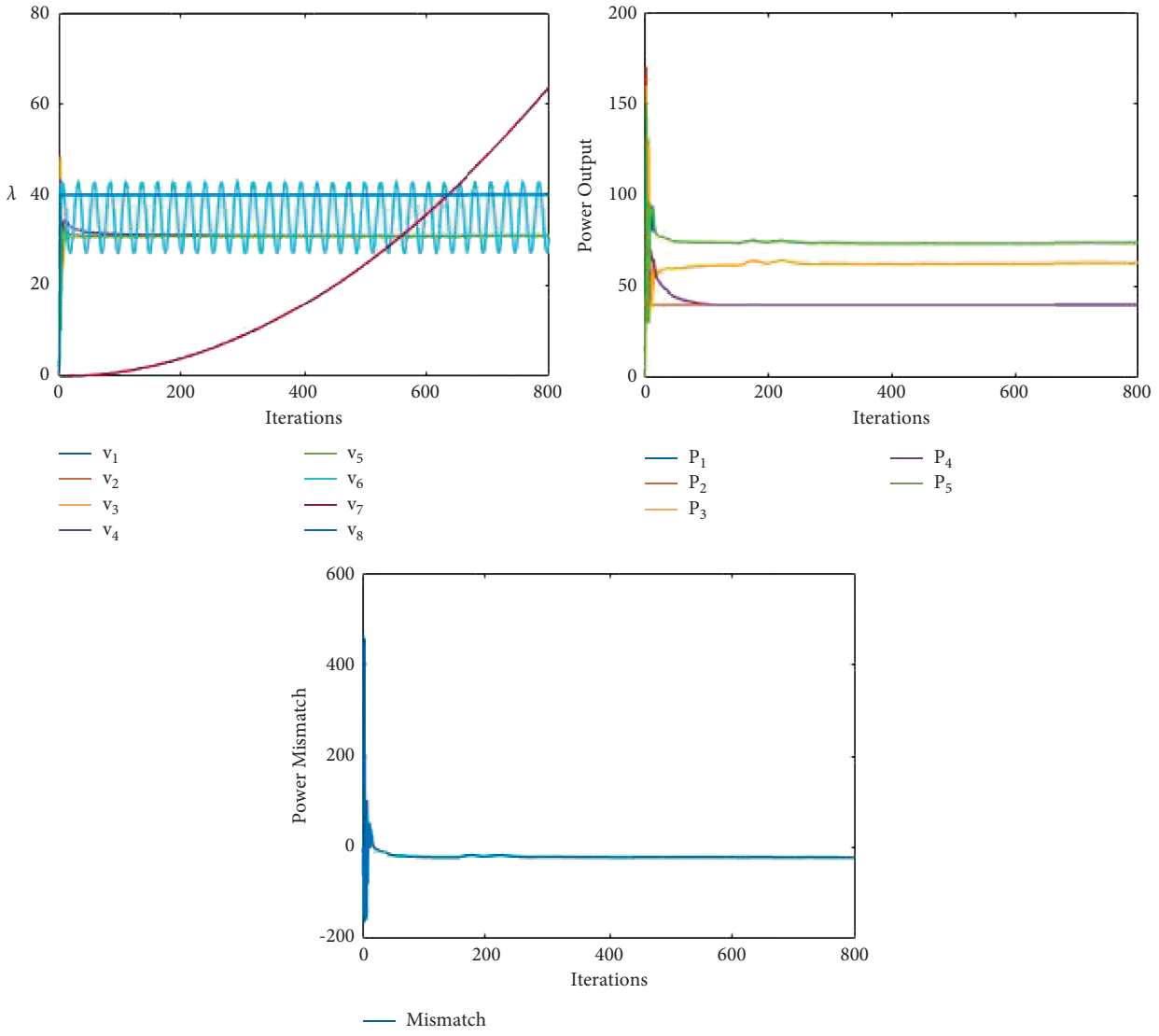


Figure 6: Variables of secure distributed energy management under FDIAs: (a) dual variables, (b) power output, and (c) power mismatch.

6. Conclusions

In this paper, a distributed energy management strategy for the port power system has been proposed under false data injection attacks. First of all, we proposed a hierarchical topology to allocate the security resources of the port power system. Then, by reconstructing the topological structure of the port information network, the robustness of the information network is improved; the impact of false data injection attacks on the port power system is reduced; and thus, the secure distributed energy management of the port energy system is realized. We have relaxed the assumption of the maximum tolerable number of attack nodes (F) while increasing the maximum number of tolerable attack nodes in the network. By protecting a portion of the nodes in the network, the normal operation of all the attacked nodes is ensured. Finally, the effectiveness of the proposed energy management was investigated by simulation results, and the defense capability of the port power system has been improved. It is important to note that our study still has some limitations. The method of topological reconstruction using the threshold changes the interactions between agents; although it is safer after topological reconstruction, there are still small fluctuations leading to energy mismatch, and it is acceptable. So the topological reconstruction algorithm still needs to be improved; This paper only studies the FDIA attacks; more types of attacks and attacks with a wider range can be considered in future work.

Nomenclature

$\tilde{p}_i$:	Power generation predicted by the renewable energy equipment i
T:	Dispatch period
$[\Delta P_i^{\min}, \Delta P_i^{\max}]$:	Confidence interval of the prediction error
a_i, b_i c_i:	Coefficients of the cost function
$P_i^{ch,\max}$:	The maximum charging power
$P_i^{ds,\max}$:	The maximum discharging power

ς^{ch}: The energy loss during the charging process
ς^{ds}: The energy loss during the discharging process
$P_i^{\min}$: The minimum power output of the device i
$P_i^{\max}$: The maximum power output of the device i
P_i^{ramp}: The ramp rate constraint
d_i: Load
$\alpha > 0$: A constant that represents the step size
$\lambda_i(k)$: The information of device i exchanged with other devices
$\zeta_i(\lambda_i(k))$: Arbitrary update function
$\chi_i^M(k)$: The maximum threshold
$\chi_i^M(k)$: The minimum threshold
$M_{n_1 \times n_1}$: The interaction between protected nodes
$M_{n_2 \times n_1}$: The interaction between protected node and ordinary nodes
$M_{n_2 \times n_2}$: The interaction of itself.

Data Availability

The data used to support this study are obtained by contacting the first author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Z. Peng, J. Wang, and D. Wang, "Distributed maneuvering of autonomous surface vehicles based on neurodynamic optimization and fuzzy approximation," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 3, pp. 1083–1090, 2018.

[2] J. Chen, T. Huang, X. Xie, P. T. W. Li, and C. Hua, "Constructing governance framework of a green and smart port," *Journal of Marine Science and Engineering*, vol. 7, no. 4, 2019.

[3] X. Lang, D. Zhongjie, and C. Jihong, "Evolutionary game of inland shipping pollution control under government co-supervision," *Marine Pollution Bulletin*, vol. 7, no. 21, 2021.

[4] L. Xu, Z. J. Di, J. H. Chen, J. Shi, and C. Yang, "Evolutionary game analysis on behavior strategies of multiple stakeholders in maritime shore power system," *Ocean & Coastal Management*, vol. 3, no. 1, 2021.

[5] Y. Li, D. W. Gao, W. Gao, H. Zhang, and J. Zhou, "Double-mode energy management for multi-energy system via distributed dynamic event-triggered Newton-raphson algorithm," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5339–5356, 2020.

[6] W. Rui, S. Qiuye, M. Dazhong, and H. Xuguang, "Line impedance cooperative stability region identification method for grid-tied inverters under weak grids," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 2856–2866, 2020.

[7] L. Ionut, "Port of Barcelona Suffers Cyberattack," 2018, https://www.bleepingcomputer.com/news/security/port-of-barcelona-suffers-cyberattack/.

[8] C. Sam, "Middle East Ports and Logistics Tech," 2020, https://m.sohu.com/a/396758697_433360.

[9] M. R. Colbie, "Cyber Attack on Asia-Pacific Ports Could Cost $110B, Hitting Global Supply Chains," 2019, https://www.insurancejournal.com/news/international/2019/10/30/546983.htm.

[10] T. Fei, S. Qihe, and L. Tieshan, "Intelligent ship integrated energy system and its distributed optimal scheduling algorithm," *ACTA AUTOMATICA SINICA*, vol. 46, pp. 1809–1817, 2020.

[11] Y. Li, H. Zhang, X. Liang, and B. Huang, "Event-triggered based distributed cooperative energy management for multienergy systems," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 14, pp. 2008–2022, 2019.

[12] L. Olatomiwa, S. Mekhilef, M. S. Ismail, and M. Moghavvemi, "Energy management strategies in hybrid renewable energy systems: a review," *Renewable and Sustainable Energy Reviews*, vol. 62, pp. 821–835, 2016.

[13] C. Deng, C. Wen, J. Huang, X.-M. Zhang, and Y. Zou, "Distributed observer-based cooperative control approach for uncertain nonlinear MASs under event-triggered communication," *IEEE Transactions on Automatic Control*, p. 1, 2021.

[14] Y. Wu and J. Dong, "Local stabilization of continuous-time T-S fuzzy systems with partly measurable premise variables and time-varying delay," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, 2018.

[15] Q. Shafiee, J. M. Guerrero, and J. C. Vasquez, "Distributed secondary control for is landed microgrids—a novel approach," *IEEE Transactions on Power Electronics*, vol. 29, no. 2, pp. 1018–1031, 2013.

[16] H. XIn, Z. Qu, J. Seuss, and M. Ali, "A self-organizing strategy for power flow control of photovoltaic generators in a distribution network," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1462–1473, 2010.

[17] G. Zhang, C. Li, D. Qi, and H. Xin, "Distributed estimation and secondary control of autonomous microgrid," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 989–998, 2016.

[18] F. Guo, C. Wen, J. Mao, J. Chen, and Y.-D. Song, "Hierarchical decentralized optimization architecture for economic dispatch: a new approach for large-scale power system," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 523–534, 2018.

[19] S. Hu, Y. Xiang, J. Liu et al., "Agent-based coordinated operation strategy for active distribution network with distributed energy resources," *IEEE Transactions on Industry Applications*, vol. 55, no. 4, pp. 3310–3320, 2019.

[20] J. Yu, C. Dou, and X. Li, "MAS-based energy management strategies for a hybrid energy generation system," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 6, pp. 3756–3764, 2016.

[21] J. Lai and X. Lu, "Resilient distributed voltage synchronization of CI networks under denial of service attacks," *IEEE Trans. Circuits Syst. II, Exp. Briefs, Nov.*vol. 24, 2020.

[22] J. Lai, X. Lu, Z. Dong, and S. Cheng, "Resilient Distributed Multiagent Control for AC Microgrid Networks Subject to Disturbances," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, 2022.

[23] X. Lu, J. Lai, and L. Guo-Ping, "Master-Slave Cooperation for Multi-DC-MGs via Variable Cyber Networks," *IEEE transactions on cybernetics*, pp. 1–14, 2021.

[24] J. H. R. van Duin, H. Geerlings, A. Verbraeck, and T. Nafde, "Cooling down: a simulation approach to reduce energy peaks of reefers at terminals," *Journal of Cleaner Production*, vol. 193, pp. 72–86, 2018.

[25] G. Parise, L. Parise, A. Malerba, F. M. Pepe, A. Honorati, and P. B. Chavdarian, "Comprehensive peak-shaving solutions for

port cranes," *IEEE Transactions on Industry Applications*, vol. 53, no. 3, pp. 1799–1806, 2017.

[26] N. Ihle, S. Runge, and N. Grundmeier, "An IT-architecture to support energy efficiency and the usage of flexible loads at a container terminal," in *Proceedings of the 28th EnviroInfo 2014 Conference*Oldenburg,Germany, September, 2014.

[27] F. D. Kanellos, "Multiagent-system-based operation scheduling of large ports' power systems with emissions limitation," *IEEE Systems Journal*, vol. 13, no. 2, pp. 1831–1840, 2019.

[28] F. D. Kanellos, E.-S. M. Volanis, and N. D. Hatziargyriou, "Power management method for large ports with multi-agent systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1259–1268, 2019.

[29] Q. Sun, N. Zhang, S. You, and J. Wang, "The dual control with consideration of security operation and economic efficiency for energy hub," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 5930–5941, 2019.

[30] Z. Liu, A. Saberi, A. A. Stoorvogel, and D. Nojavanzadeh, "H∞ almost state synchronization for homogeneous networks of non-introspective agents: a scale-free protocol design," *Automatica*, vol. 122, Article ID 109276, 2020.

[31] J. Zhang, X. Chen, and G. Gu, "State Consensus for Discrete-Time Multi-Agent Systems over Time-Varying Graphs," *IEEE Transactions on Automatic Control*, vol. 66, 2020.

[32] Q. Guo, S. Xin, J. Wang, and H. Sun, "Comprehensive security assessment of information energy system from the blackout in Ukraine," *Power system automation*, vol. 40, no. 05, pp. 145–147, 2016, in Chinese.

[33] Y. Li, D. W. Gao, W. Gao, H. Zhang, and J. Zhou, "A Distributed Double-Newton Descent Algorithm for Cooperative Energy Management of Multiple Energy Bodies in Energy Internet," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, 2020.

[34] B. Huang, Y. Li, F. Zhan, Q. Sun, and H. Zhang, "A distributed robust economic dispatch strategy for integrated energy system considering cyber-attacks," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 880–890, 2022.

[35] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber-physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, 2012.

[36] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, "Cyber Security Analysis of State Estimators in Electric Power Systems," in *Proceedings of the IEEE 49th Conf. Decis. Control (CDC)*, pp. 5991–5998, Atlanta, GA, USA, December, 2010.

[37] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 645–658, 2011.

[38] J. Zhang, H. Zhang, and T. Feng, "Distributed optimal consensus control for nonlinear multiagent system with unknown dynamic," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3339–3348, 2018.

[39] Y. Tang, C. Qian, M. Li, and Q. Wang, "Overview on cyber-attacks against cyber physical power system," *Automation of Electric Power Systems*, vol. 40, no. 17, pp. 59–69, 2016, in Chinese.

[40] M. R. Habibi, H. R. Baghaee, T. Dragicevic, and F. Blaabjerg, "False data injection cyber-attacks mitigation in parallel DC/DC converters based on artificial neural networks," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 2, pp. 717–721, 2021.

[41] A. A. Khan, O. A. Beg, and M. Alamaniotis, "Intelligent anomaly identification in cyber-physical inverter-based systems," *Electric Power Systems Research*, vol. 193, 2021.

[42] S. Sahoo, J. C.-H. Peng, A. Devakumar, S. Mishra, and T. Dragicevic, "On detection of false data in cooperative DC microgrids-A discordant element approach," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 8, pp. 6562–6571, 2020.

[43] S. Kar, S. R. Samantaray, and M. D. Zadeh, "Data-mining model based intelligent differential microgrid protection scheme," *IEEE Systems Journal*, vol. 11, no. 2, pp. 1161–1169, 2017.

[44] E. Casagrande, W. L. Woon, H. H. Zeineldin, and D. Svetinovic, "A differential sequence component protection scheme for microgrids with inverter-based distributed generators," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 29–37, 2014.

[45] Y. Liu, H. Xin, Z. Qu, and D. Gan, "An attack-resilient cooperative control strategy of multiple distributed generators in distribution networks," *IEEE Transactions on Smart Grid*, vol. 7, no. 6, pp. 2923–2932, 2016.

[46] S. Abhinav, H. Modares, F. L. Lewis, and A. Davoudi, "Resilient cooperative control of DC microgrids," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 1083–1085, 2019.

[47] S. Abhinav, H. Modares, F. L. Lewis, F. Ferrese, and A. Davoudi, "Synchrony in networked microgrids under attacks," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6731–6741, 2018.

[48] H. J. LeBlanc, H. Zhang, X. Koutsoukos, and S. Sundaram, "Resilient asymptotic consensus in robust networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 4, pp. 766–781, 2013.

[49] W. Abbas, Y. Vorobeychik, and X. Koutsoukos, "Resilient consensus protocol in the presence of trusted node," *Proc. 7th Int. Symp. Resilient Control Syst.*, pp. 1–7, 2014.

[50] C. Zhao, J. He, and Q.-G. Wang, "Resilient distributed optimization algorithm against adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 65, no. 10, pp. 4308–4315, 2020.

[51] M. Acciaro, T. Vanelslander, C. Sys et al., "Environmental sustainability in seaports: a framework for successful innovation," *Maritime Policy & Management*, vol. 41, no. 5, pp. 480–500, 2014.

[52] G. Harry and V. Duin Ron, "A new method for assessing CO2-emissions from container terminals: a promising approach applied in Rotterdam," *Journal of Cleaner Production*, vol. 19, no. 6-7, pp. 657–666, 2011.

[53] D. Ma, X. Hu, H. Zhang, Q. Sun, and X. Xie, "A hierarchical event detection method based on spectral theory of multi-dimensional matrix for power system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 4, pp. 2173–2186, 2021.

[54] K. Ma, Y. Q. Yu, S. Y. Zhu, J. Yang, and X. Guan, "Distributed algorithm for economic dispatch based on gradient descent and consensus in power grid," *Sci Sin Inform*, vol. 48, pp. 1364–1380, 2018, in Chinese.

[55] S. Mei, A. Xue, and X. Weng, "Summary on Risk of Blackouts in Complex Interconnected Power Grids and Prospects of its Preventive control," in *Proceedings of the the 24th Chinese Control Conference*, China Automation Society, Guangzhou, China, August, 2006, in Chinese.

[56] L. Yao, N. Peng, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security*, vol. 14, no. 1, pp. 1–33, 2011.

[57] Q. Sun, R. Fan, Y. Li, B. Huang, and D. Ma, "A distributed double-consensus algorithm for residential we-energy," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4830–4842, 2019.

Hindawi
Complexity
Volume 2021, Article ID 3434148, 14 pages
https://doi.org/10.1155/2021/3434148

Research Article

Reliability Evaluation for Cyber-Physical Smart Substation

Yangrong Chen,[1] **Jun'e Li**,[1] **Ang Xu**,[2] **Kai Yuan**,[3] **and Kaipei Liu**[3]

[1]*Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China*
[2]*State Grid Wuhan Electric Power Supply Co., Ltd, Wuhan 430010, China*
[3]*School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China*

Correspondence should be addressed to Jun'e Li; jeli@whu.edu.cn

Received 29 April 2021; Revised 6 July 2021; Accepted 7 October 2021; Published 21 October 2021

Academic Editor: Qiuye Sun

Smart substation is the key part of smart grid. The reliability of smart substation is extremely important to the safe and stable operation of the smart grid. Smart substation is a cyber-physical system (CPS). Hence, this paper conducts reliability evaluation study for smart substation from the perspective of CPS. Firstly, the basic reliability indices of cyber and physical elements of smart substation are presented. The reliability index of one cyber element takes into account the reliability factors of data leakage, tampering, loss and delay, etc., on the cyber side. Then, the cyber-physical interactions of smart substation are analyzed. It is concluded that the effect of the cyber side and cyber-physical interactions on the reliability of smart substation is reflected in the effect of measurement and control messages on circuit breaker operation. And, the new reliability indices considering cyber-physical interactions are proposed. Furthermore, the MALI-hybrid method, which combines the Monte Carlo method, analytical method, Latin hypercube sampling method, and important sampling method, is presented for evaluating the reliability of smart substation. Finally, the rationality of the proposed reliability indices, the efficiency, and correctness of MALI-hybrid method are verified by case studies.

1. Introduction

With the increasing intelligent electronic devices (IED) installation, smart substation can be regarded as a typical cyber-physical systems (CPSs), in which various types of information are being exchanged on a communication network [1]. Interruption, delay, communication errors, and various cyber attacks on the cyber side of smart substation probably cause misoperation or refusal of circuit breakers (CBs) and even affect the safe and stable operation of smart substation [2–6]. Meanwhile, malfunction on the physical side of smart substation may make cyber system lose right perception and real-time control for physical system, which in turn further affects the operation of power system and forms cascading failures [7, 8]. Thus, reliability evaluation for smart substation should consider the reliability factors on both the physical side and cyber side, and the interactions between the two sides.

1.1. Motivation. At present, there are few studies on the reliability evaluation of smart substation or power grid from the perspective of CPSs. Zhang [9] constructed a Markov model of physical elements under the influence of the secondary system and used state-space method to calculate the reliability of the physical elements considering the influence of the secondary system. However, the effect of the primary system on the reliability of the secondary system and the interactions between the two systems in smart substation are not considered. To evaluate the effect of communication system on power system, Tang et al. [10] established dynamic vulnerability matrix for evaluating real-time vulnerability of composite system considering communication faults such as time delay, bit error, and interruption. However, the effect of the power system on the communication system is not considered, and the research is not targeted at smart substation. Lei et al. [11] proposed a novel reliability modelling and analysis methodology using

cyber-physical interface matrix for modern substation protection systems. The study provides an idea for evaluating the reliability of substation based on cyber-physical integration, but research about reliability index system and quantitative calculations method is insufficient. Ten et al. [12] used existing power system simulators to evaluate the reliability of smart substations considering cyber attacks. However, this research focuses on analyzing the effect of cyber attacks on substations and does not consider the effect of the physical side on the cyber side and cyber-physical interactions on the reliability of substation. Liu et al. [13] presented an analytical method to quantify the effect of cyber fault on the reliability of physical system during distribution automation considering dynamic routing, delay, and communication error, particularly the cyber traffic. However, the study disregards the effect of the physical side on the cyber side and interactions between them. Aravinthan et al. [14] summarized the status and needs for reliability evaluation of cyber-physical power systems (CPPS) and standardized the CPPS model for reliability computation. This model could be a reference for research of reliability evaluation but cannot be directly used to evaluate the reliability of smart substation.

In addition, as we all know, stability is an important condition for maintaining the safe and reliable operation of the power system. The study about stability can help us do a better reliability evaluation. Wang et al. [15] proposed a droop coefficients stability region analysis approach for power system. Sun et al. [16] studied the stability analysis issues of Energy Internet. The authors in [17] have invented a novel small-signal modelling method based on the characteristic equation to assess the stability of complex microgrids. reference [18] proposed a line impedance cooperative stability region identification method for grid-tied inverters under weak grids. Hu et al. [19] pointed that open-end winding machines provide better controllability and reliability than conventional start-connected machines. The above work provides guidance for power system reliability evaluation method researches.

Further, the data leakage, tampering, loss, network interruption, and delay also occur frequently in other industrial systems (e.g., pipeline network systems) [20]. And with the development of pipeline transportation, the requirements for safety and reliability of pipeline network systems have increased in recent years [21]. Thus, the study of new reliability indices considering cyber-physical interactions and evaluation methods can also provide the reference for other industrial systems.

From the above analysis, it can be seen that there are few literatures about reliability analysis and reliability evaluation method for smart substation from the perspective of CPSs. In our recent study, the interaction mechanism of the cyber side and the physical side of smart substation were analyzed preliminarily [22]. However, the reliability evaluation method is relatively simple and has low accuracy, the case study has some deficiencies, and the expression of formulas is not easy to read and understand for readers. Based on the above review of reliability evaluation and Ref. [22], this paper fully analyzes various reliability factors and how they affect the reliability of smart substation. Then, the calculation expressions of reliability indices are modified and improved so as to become easy to read and understand for readers. Furthermore, the MALI-hybrid reliability evaluation method, which has fast simulation speed and high precision, is presented. Finally, a more practical case is presented to verify the rationality of the proposed reliability indices and reliability evaluation method. Simulation results show that the proposed reliability indices are reasonable, and the presented reliability evaluation method can save simulation time with improving accuracy.

1.2. Contribution. The main contributions of this paper can be summarized as follows:

(1) A thorough analysis on the interactions between the cyber side and the physical side of smart substation is conducted. It is concluded that the reliability evaluation of smart substation should consider the effect of power grid disturbance and complicated operation states on the correctness of measurement information, and the effect of all factors on the reliability of smart substation is carried out by measurement and control messages, which affect the power system's operation by the CB's misoperation or refusal.

(2) The reliability indices are proposed for smart substation considering the reliability factors of the cyber side, the physical side, and the cyber-physical interactions. These indices involve the reliability index of cyber equipment considering its importance, the reliability index of communication links considering the reliability of data transmitting on them, and the equivalent reliability indices of CB considering cyber-physical interactions. The importance of cyber equipment refers to its effect on the communication business when the data leakage, packet error, message tampering or forgery, packet loss or interruption, and message delay occur, and it can be calculated by the proposed AHP model.

(3) A MALI-hybrid reliability evaluation method, which combines the Monte Carlo method, analytical method, LHS method, and important sampling method, is presented. This method has faster simulation speed and more precise evaluation results than the common reliability evaluation methods.

1.3. Organization. The rest of this paper is organized as follows. Section 2 presents the elements for reliability evaluation of smart substation. Section 3 gives the basic reliability indices of smart substation. Section 4 analyzes the interactions between the cyber side and the physical side of smart substation and proposes the reliability indices considering the effect of cyber-physical interactions. Section 5 presents a MALI-hybrid reliability evaluation method for smart substation. Section 6 verifies the rationality of the proposed reliability indices and the reliability evaluation method. Section 7 gives a summary and a future outlook.

2. Elements for Reliability Evaluation of Smart Substation

The D2-1 substation defined in IEC 61850 standard is widely used in practical engineering construction. Thus, we choose D2-1 smart substation as analysis object and elaborate the proposed reliability indices and reliability evaluation method. For better understanding the interactions between the cyber side and the physical side of smart substation, the communication network of D2-1 substation, of which the architecture is "three layers, two networks" as a common one of the present smart substations, is shown in Figure 1.

The topology of D2-1 smart substation is shown in Figure 2 [23]. The smart substation in Figure 2 contains 1 bus interval (BI), 6 feeder intervals (FI1-FI6), and 2 transformer intervals (TI1-TI2). The physical side contains buses, CBs, DSs, electrical voltage transformers (EVTs), and electrical current transformers (ECTs), which are connected by electrical lines. The cyber side contains merging unit (MU), intelligent electronic devices (IEDs), protection and control (P&C) IEDs, CB IEDs, and switches, which are connected by communication links.

Reliability evaluation of smart substation should consider the reliability of all the above elements. For formal description, the above elements are represented by nodes and branches as follows.

The MU IEDs, P&C IEDs, CB IEDs, and switches on the cyber side are abstracted as nodes of cyber system, which can be defined as the set N_c:

$$N_c = \left\{N_{c_1}, N_{c_2}, \ldots, N_{c_g}\right\}, \tag{1}$$

where the set $\left\{N_{c_1}, N_{c_2}, \ldots, N_{c_g}\right\}$ is the cyber equipment set on the cyber side.

The communication links are abstracted as branches of cyber system, which are defined as the set B_c:

$$B_c = \left\{B_{c_1}, B_{c_2}, \ldots, B_{c_k}\right\}, \tag{2}$$

where the set $\left\{B_{c_1}, B_{c_2}, \ldots, B_{c_k}\right\}$ is the communication link set on the cyber side.

The buses, transformers, CBs, DSs, EVTs, and ECTs on the physical side are abstracted as nodes of physical system, which can be defined as the set N_p:

$$N_p = \left\{N_{p_1}, N_{p_2}, \ldots, N_{p_m}\right\}, \tag{3}$$

where the set $\left\{N_{p_1}, N_{p_2}, \ldots, N_{p_m}\right\}$ is the physical equipment set on the physical side.

In smart substation, though the connection wires between the physical devices are short, they are still abstracted as branches of physical system and are called primary electrical lines in this paper for the sake of generality. The connection wires are defined as the set B_p:

$$B_p = \left\{B_{p_1}, B_{p_2}, \ldots, B_{p_n}\right\}, \tag{4}$$

where the set $\left\{B_{p_1}, B_{p_2}, \ldots, B_{p_n}\right\}$ is the primary electrical line set on the physical side.

3. The Basic Reliability Indices

To study the factors affecting the reliability of smart substation, this section analyzes the basic reliability indices of elements in smart substation.

Reliability refers to the possibility of the equipment to operate normally under specified conditions over a period. Traditionally, the reliability of one element can be obtained according to the following equation:

$$R = \frac{u}{\lambda + u}, \tag{5}$$

where λ and u are the failure rate and repair rate of the element, respectively.

3.1. Reliability Indices of Cyber Elements. As described in Section 2, cyber elements contain cyber equipment and communication links.

3.1.1. Reliability Indices of Cyber Equipment. The reliability factors of the cyber side play on the physical side through communication business. And the reliability of each communication device on the transmission link has different effects on communication business. Therefore, we use the equivalent failure rate λ_{e_i} of cyber equipment to take the place of λ in equation (5). λ_{e_i} can be defined as follows:

$$\lambda_{e_i} = \partial_i \cdot \lambda_i, \tag{6}$$

where ∂_i is the importance degree of the cyber equipment (i.e., the effect of equipment failure on the reliable transmission of communication business; it is valued between 0 and 1); λ_i is the failure rate of cyber equipment, the same as the meaning of λ in equation (5).

In this paper, we employ an analytic hierarchy process (AHP) method to calculate the importance degree ∂_i of cyber equipment. The AHP model is shown in Figure 3.

In Figure 3, the failure rate of cyber equipment is taken as the index layer. The effect of cyber equipment failure on the security, correctness, and real-time of communication business is taken as the criterion layer. The comprehensive effect of cyber equipment failure on communication business is taken as the target layer. In the criterion layer, the effect of cyber equipment failure on the security of communication business refers to data leakage caused by equipment failure; the effect of cyber equipment failure on the correctness of communication business refers to packet error, tampering, and forgery of service messages caused by equipment failure; the effect of cyber equipment failure on the real-time of communication business refers to delay, loss, and interruption of service messages caused by equipment failure. It can be seen that the model takes into account both conventional faults and failures caused by cyber attacks. For the detailed calculation process of importance degree ∂_i, refer to Refs. [24].

3.1.2. Reliability Indices of Communication Links. The reliability of the communication link is not only related to the

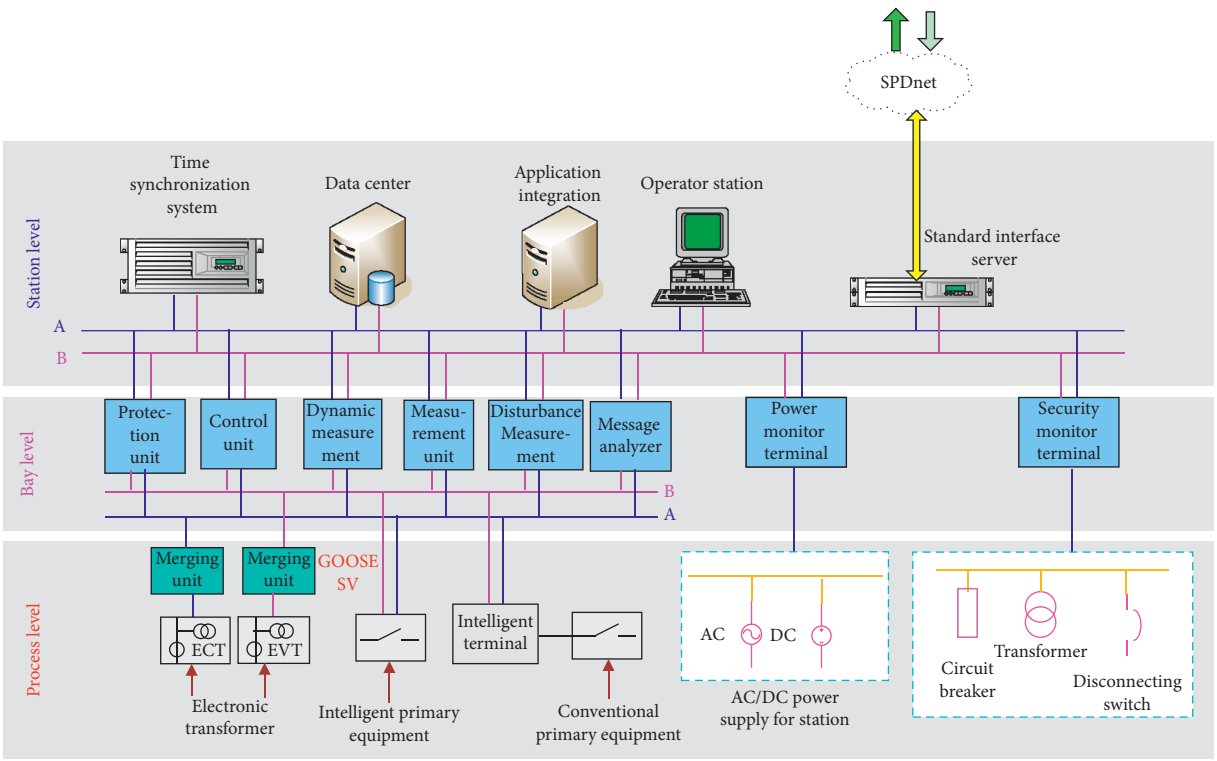


Figure 1: The architecture of D2-1 smart substation.

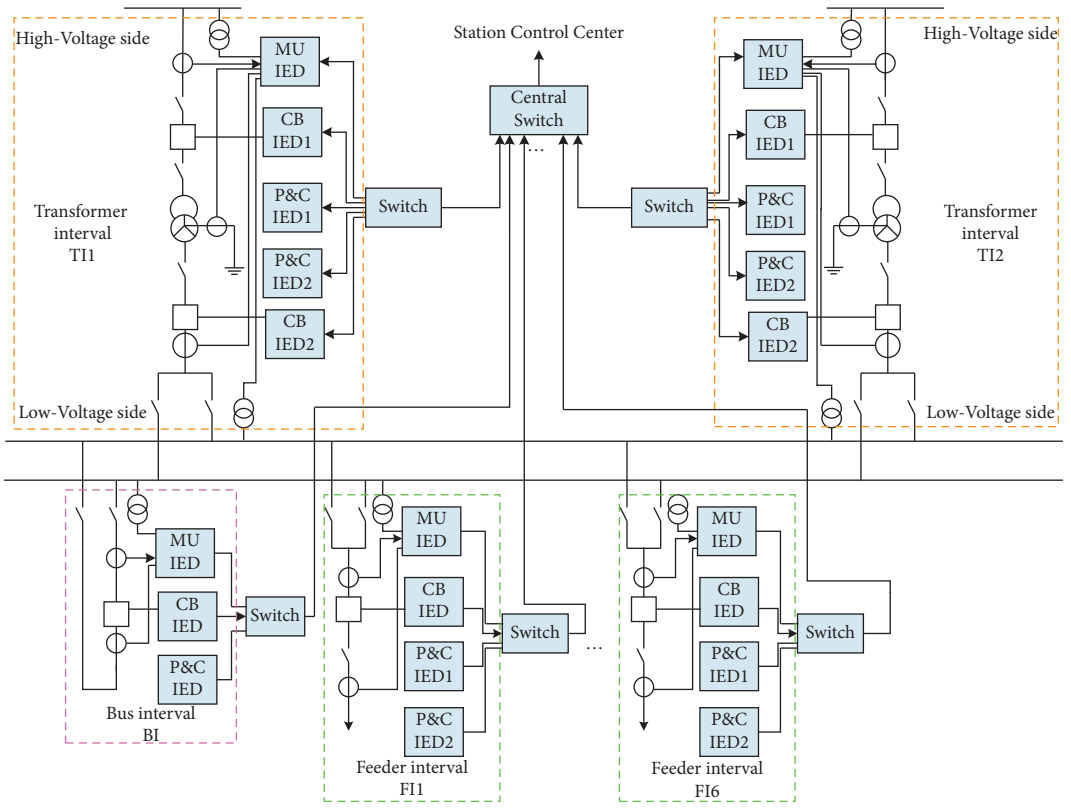


Figure 2: Topology diagram of D2-1 smart substation.

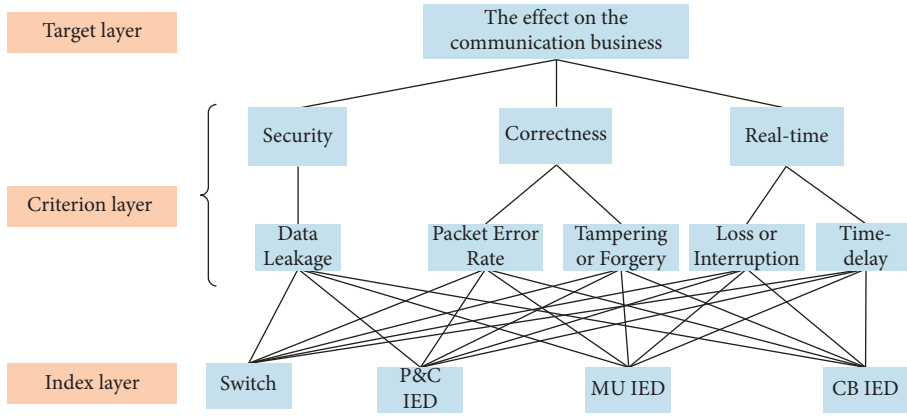


Figure 3: AHP model for importance degree evaluation of cyber equipment.

reliability of the link itself, but also related to the reliability of data transmitting on it. According to the reliability calculation method of cascading systems, the reliability of one communication link can be calculated as follows:

$$R\left(B_{c_k}\right) = R\left(B_{c_{i\longrightarrow j}}\right) = CL_{i\longrightarrow j} \cdot D, \tag{7}$$

where $R(B_{c_k})$ is the reliability of the kth communication link, a communication link from the cyber equipment N_{c_i} to N_{c_j}, so it is also noted as $R(B_{c_{i\longrightarrow j}})$; $CL_{i\longrightarrow j}$ is the reliability of the link hardware, and D is the reliability of data transmitting on the link; $CL_{i\longrightarrow j}$ can be calculated by equation (5).

The reliability of data transmitting on the link is mainly related to propagation delay and error. In smart substation, optical fibers are used as the communication links, and the links are short. In the case of no attack, the delay is relatively stable and extremely small, so it can be ignored. Packets in smart substation have error detection mechanism, but any error detection mechanism has the possibility of missing detection. Therefore, this paper uses packet error rate (i.e., the omission rate of error detection method of the packet) to represent the effect of transmitting error on the reliability of data. We define P_e as the packet error rate. Thus, D in equation (7) can be calculated as follows:

$$D = 1 - P_e. \tag{8}$$

3.2. Reliability Indices of Physical Elements. The reliability indices of the physical equipment and the primary electrical lines are usually obtained according to the operation of smart substation. This paper takes availability as the basic reliability index of physical elements, and the reliability can be calculated by equation (5).

4. Cyber-Physical Interactions and the Reliability Indices

The basic indices in Section 3 only consider the reliability factors of a single side, but do not consider the effect from another side and the interactions. This section analyzes the interactions between the cyber side and the physical side of smart substation, and then the reliability indices of smart substation considering the effect of the cyber side, the effect of the physical side, and the interactions are presented.

4.1. Interactions between Cyber Side and Physical Side. The main function of smart substation is to convert high and low voltage. The bus voltage monitoring and tripping, two of the communication business of smart substation, can best reflect the interactions between cyber side and physical side. Hence, we take the bus voltage monitoring and tripping as the example to analyze the cyber-physical interaction mechanism of smart substation.

The process of bus voltage monitoring and tripping can usually be described as follows. First, bus voltage from the physical side is sampled and converted into analog or digital signals by local measurement equipment. Second, the data enclosed in packets is transmitted to the control center of the station level via the communication networks. Third, a centralized application conducts online computation on the basis of the received packets and generates tripping commands to regulate physical equipment. Forth, these control commands are sent back to the physical side via the communication networks again, and the related physical devices take actions according to the tripping commands they received. Finally, the physical state of the smart substation is changed as a consequence of the closed-loop control. Both the beginning and end of the closed-loop control are connected to the physical side; thus, the cyber side and the physical side in smart substation show an interactive relationship [25–27].

4.1.1. Effect of the Cyber Side on the Physical Side. Measurement information (e.g., bus voltage) from the physical side transmits from EVTs/ECTs to MU IEDs, and then the measurement information enclosed in sampled value (SV) messages is relayed to P&C IEDs. P&C IEDs decide whether to send the control messages according to SV messages and control strategy. Finally, P&C IEDs send generic object oriented substation event (GOOSE) messages containing tripping messages to the corresponding CB IEDs and control the action of CBs.

During the above communication process, cyber contingencies, no matter whether they are accidental faults or malicious attacks, may lead to the inappropriate operation of CBs, even a large-scale power outage. Moreover, the above description shows that the effect of the cyber side on the physical side in smart substation is mainly conducted through controlling CBs. And the control process is carried out by communication messages and equipment of the cyber side.

4.1.2. Effect of the Physical Side on the Cyber Side and the Cyber-Physical Interactions. The effect of the physical side on the cyber side and the cyber-physical interactions of smart substation is as follows.

Physical equipment supplies power for local cyber equipment. Failure of physical system will result in the inability to supply power for cyber equipment [28]. While in practice, an independent uninterrupted power supply is usually used to supply power for cyber equipment when the power system fails. That is, a short power outage does not affect the normal operation of cyber system.

When the power system fails, it is very likely that multiple IEDs will send out a large number of GOOSE messages in a short interval [29]. The surge of GOOSE messages might cause the network congestion and messages to be delayed and even cause system decision-making error. The congestion caused by power system fails, excluding attacks, can be avoided by a well-planned network of the cyber system.

ETs are more susceptible to electromagnetic interference in operation than traditional transformers due to their diverse sensing principles and a large number of electronic components in the primary body [7]. Specially, since the installation site of ETs is close to the electrical lines, they are more susceptible to interference through direct conduction and electromagnetic field coupling under the conditions of on-off operation and short-circuit. The effect of electromagnetic interference on ETs may lead to errors in measurement information, and even malfunction of ETs, which may make protection or regulation go awry.

Because of the complex structure of the power grid, there is sympathetic interaction between the nonlinear ferromagnetic elements, such as the transformers and the ETs, which might result in the complex transient process [30]. This might lead the existing detection methods to correctly identify the operation state of power grid, then the protective devices malfunction due to error of measurement information.

Due to the access of new energy systems, the application of AC/DC (alternating current/direct current) hybrid transmission, and the application of FACTS (flexible alternating current transmission) devices, the behavior of modern power systems is becoming more and more complicated. Meanwhile, most of the existing protective devices filter out high-frequency transient components and do not use high-frequency transient components. Therefore, for the AC/DC hybrid power systems connected with new energies, the detection method that only retains power frequency components and finite harmonic components cannot accurately detect the operating status of complicated power systems and may give wrong measurement information and lead to wrong protection actions.

The reliable operation of smart substation relies on accurate detection and identification to power system status (e.g., operating status, fault identification, etc.). According to the above analysis, we can conclude that (1) the measurement information errors caused by grid disturbance and complicated running state become the main factors to be considered in the reliability evaluation of smart substation; (2) in smart substation, both the effect of cyber side on physical side, the effect of physical side on cyber side, and the interaction are delivered through the measurement and control messages, and the misoperation or refusal of the CBs resulting from wrong measurement or control messages will further affect the safe and reliable operation of the power system. Thus, the effect of the cyber side and cyber-physical interactions on the reliability of smart substation can be obtained through analyzing the effect of measurement and control messages on the actions of CBs.

4.2. Message Flows Related to Cyber-Physical Interactions. To analyze the reliability of measurement and control messages and their effect on the action of CBs, it is needed to analyze the transmission paths of measurement and control messages. The reliability of the messages can be calculated according to the reliability of the elements on the transmission path. Further, equivalent reliability of CBs considering the cyber-physical interactions is obtained.

In smart substation, there are SV messages, GOOSE messages, manufacturing message specification (MMS) messages, etc. According to the analysis in Section 4.1, the interaction between the cyber side and the physical side is mainly related to the process level. Therefore, to simplify the analysis, this paper mainly focuses on analyzing the SV and GOOSE messages of the process level. The simplified communication network topology of the smart substation in Figure 1 and the message flows of its process level are shown in Figure 4. In this figure, only one transformer interval and one feeder interval are presented for simplifying the description, but all of the elements in smart substation should be involved in practice.

In Figure 4, the digit on each device is the node number of the device. S1, S2, . . ., S5 refer to the flows of SV messages encapsulating the measurement information (e.g., voltage, current, and system frequency). G1, G2, . . ., G14 refer to the GOOSE message flows, in which G1, G2, . . ., G4 are the flows of messages encapsulating breaker positions, G5, G6, . . ., G9 are the flows of messages encapsulating remote control commands, G10, G11 and G12 are the flows of messages encapsulating primary remote information, and G13 and G14 are the flows of messages encapsulating blocking signals.

We denote the transmission path as L_{i-j}, referring to the set of nodes on the transmitting path of one message, in which i is the number of the source nodes and j is the number of the destination nodes of the message. Table 1

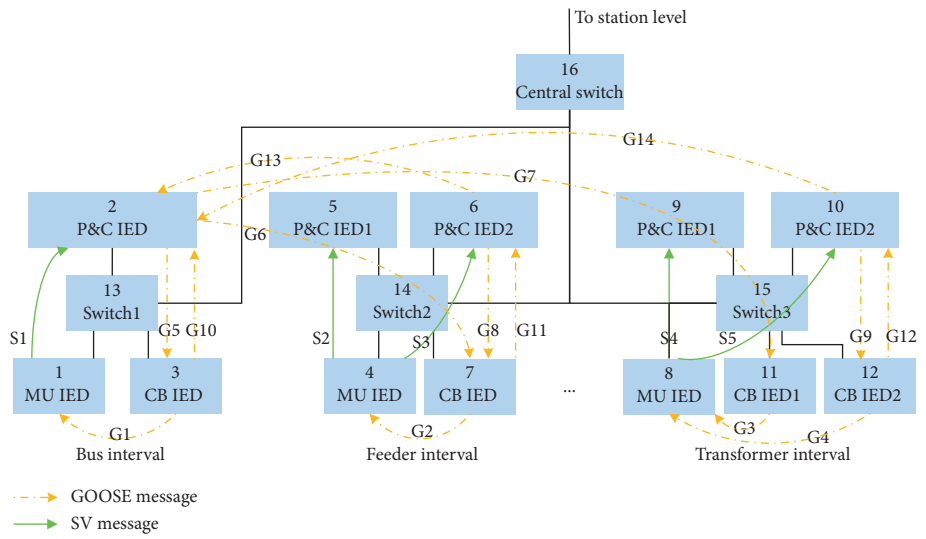


Figure 4: Communication network topology and message flows of the process level in smart substation.

Table 1: Transmission paths of the message flows in Figure 4.

Message flow	Denotation	Transmission path
S1	L_{1-2}	$1 \longrightarrow 13 \longrightarrow 2$
S2	L_{4-5}	$4 \longrightarrow 14 \longrightarrow 5$
S3	L_{4-6}	$4 \longrightarrow 14 \longrightarrow 6$
S4	L_{8-9}	$8 \longrightarrow 15 \longrightarrow 9$
S5	L_{8-10}	$8 \longrightarrow 15 \longrightarrow 10$
G1	L_{3-1}	$3 \longrightarrow 13 \longrightarrow 1$
G2	L_{7-4}	$7 \longrightarrow 14 \longrightarrow 4$
G3	L_{11-8}	$11 \longrightarrow 15 \longrightarrow 8$
G4	L_{12-8}	$12 \longrightarrow 15 \longrightarrow 8$
G5	L_{2-3}	$2 \longrightarrow 13 \longrightarrow 3$
G6	L_{2-7}	$2 \longrightarrow 13 \longrightarrow 16 \longrightarrow 14 \longrightarrow 7$
G7	L_{2-11}	$2 \longrightarrow 13 \longrightarrow 16 \longrightarrow 15 \longrightarrow 11$
G8	L_{6-7}	$6 \longrightarrow 14 \longrightarrow 7$
G9	L_{10-12}	$10 \longrightarrow 15 \longrightarrow 12$
G10	L_{3-2}	$3 \longrightarrow 13 \longrightarrow 2$
G11	L_{7-6}	$7 \longrightarrow 14 \longrightarrow 6$
G12	L_{12-10}	$12 \longrightarrow 15 \longrightarrow 10$
G13	L_{6-2}	$6 \longrightarrow 14 \longrightarrow 16 \longrightarrow 13 \longrightarrow 2$
G14	L_{10-2}	$10 \longrightarrow 15 \longrightarrow 16 \longrightarrow 13 \longrightarrow 2$

presents the transmission paths of the message flows in Figure 4.

4.3. Reliability Indices considering Cyber-Physical Interactions. The nodes of one transmission path are in series. Thus, the reliability of one message can be obtained according to the following equation:

$$R_{\text{message}} = R\left(N_{c_1}\right) \cdot R\left(N_{c_2}\right) \cdot \ldots \cdot R\left(N_{c_i}\right) \cdot R\left(B_{c_1}\right) \cdot R\left(B_{c_2}\right) \cdot \ldots \cdot R\left(B_{c_k}\right), \tag{9}$$

where $R(N_{c_i})$ is the reliability of cyber equipment, the same as the meaning of R in equation (5); $R(B_{c_k})$ is the reliability of communication link, the same as the meaning of $R(B_{c_k})$ and $R(B_{c_{i\longrightarrow j}})$ in equation (7).

4.3.1. Reliability of CB considering the Reliability Factors of the Cyber Side. As concluded above, the cyber side affects the physical side through GOOSE messages controlling CBs and affected by SV messages in smart substation. Therefore, the reliability of one CB, considering the reliability factors of the cyber side, can be presented in the following equation:

$$R'\left(N_{p_{\text{CB}}}\right) = R\left(N_{p_{\text{CB}}}\right) \cdot R_{\text{SV}} \cdot R_{\text{GOOSE}}, \tag{10}$$

where $R(N_{p_{\text{CB}}})$ is the reliability of one CB without considering the reliability factors of the cyber side; R_{SV} is the reliability of the SV messages without considering the measurement information errors; R_{GOOSE} is the reliability of the GOOSE messages; R_{SV} and R_{GOOSE} can be calculated according to equation (9).

4.3.2. Reliability of CB considering the Effect of the Physical Side on the Cyber Side and Cyber-Physical Interactions. The effect of the physical side on the cyber side is mainly caused by measurement information and its transmission. The reliability of SV messages considering the effect of the physical side on the cyber side can be defined as

$$R_{\mathrm{SV}}' = (1 - \varsigma_r) \cdot R\left(N_{p_{\mathrm{ET}}}\right) \cdot R_{\mathrm{SV}}, \tag{11}$$

where $R(N_{p_{\mathrm{ET}}})$ is the reliability of one ET; ς_r is the error rate of measurement information, which refers to the error occurrence probability of the measurement information transmitted to MU when the ET operates normally. The error rate is mainly affected by grid disturbance and complicated running state and can be obtained from actual statistics of running data. R_{SV} is the same as that in equation (10).

When the reliability of SV messages is reduced, the corresponding CBs cannot be controlled correctly by the cyber side in smart substation. That is, the effect of physical side on cyber side might be ultimately revealed on the physical side of smart substation. Therefore, considering the interactions between the cyber side and the physical side, the reliability of one CB can be defined as

$$R''\left(N_{p_{\mathrm{CB}}}\right) = R\left(N_{p_{\mathrm{CB}}}\right) \cdot R_{\mathrm{SV}}' \cdot R_{\mathrm{GOOSE}}, \tag{12}$$

where $R(N_{p_{\mathrm{CB}}})$ and R_{GOOSE} are the same as those in equation (10), and R_{SV}' is the same as that in equation (11).

5. Reliability Evaluation Method for Smart Substation

We take D2-1 smart substation shown in Figure 2 as an example to illustrate the reliability evaluation method proposed in this paper.

5.1. System Reliability Index of Smart Substation. The main electrical connection can reflect the connection relationship of primary equipment in smart substation. The main electrical connection of D2-1 smart substation is shown in Figure 5.

To analyze the reliability of smart substation, we convert the main electrical connection diagram in Figure 5 into a logical correlation diagram as shown in Figure 6. To simplify the simulation and calculation, only two feeder lines are presented in Figure 6.

The circle in Figure 6 represents the combination of the primary equipment. L1 and L2 represent the series combination of transformer, electrical lines, EVT, buses, and DSs; L5 and L6 represent the series combination of buses, ECT, electrical lines, and DSs; L9 and L10 represent the series combination of transformer, electrical lines, and DSs; L3, L4, L7, L8, and L11 represent CBs. The reliability of each circle can be obtained based on the reliability calculation equation $R_S = \prod R_i$ of Cascading Systems.

In the remaining sections, we define the probability P_S that the main electrical connection operates normally as the reliability index of smart substation. P_S is defined as follows:

$$P_S = \sum_{i \in S} p_i, \tag{13}$$

where S is the set of all running states; p_i is the probability of smart substation working in the reliable state (i.e., Load1 and Load2 are supplied power normally), and p_i can be obtained according to the following reliability analysis method.

5.2. Reliability Evaluation Method

5.2.1. Analysis of Existing Reliability Evaluation Methods. The common reliability evaluation methods are generally divided into analytical method and simulation method [31].

The analytical method is used for evaluating the reliability of smart substation through mathematical modelling. Lei et al. [11] used the cyber-physical interface matrix and analytical method to evaluate the effect of each component failure on load shedding in IEC 61850 based substation. Falahati et al. [25] used the analytical method to evaluate the reliability of smart substations and microgrid CPS with simple structure. However, obtaining high-precision results depends on some previous assumptions [32].

The simulation method mainly refers to the Monte Carlo method. Liu et al. [33] used the nonsequential Monte Carlo method for information domain and physical domain simulation, respectively, when evaluating the reliability of the active distribution network. Wang et al. [34] employed the nonsequential Monte Carlo method to simulate the operation of the microgrid CPS. However, the nonsequential Monte Carlo method does not consider the temporal properties of components in practice. Sun and Xie [35] used the sequential Monte Carlo method to describe the correlation between the time series of wind speeds and loads and evaluated the reliability of distribution systems containing wind turbine generator. Araújo et al. [36] used sequential Monte Carlo method to select the system scenarios resulting in uncertainties associated with load fluctuations, load forecasting errors, distributed generation unavailability, and intermittence of renewable energy resources. Zhou et al. [37] proposed a sequential Monte Carlo simulation method based on load path. The method can be used for evaluating the reliability of the power grid through traversal search for load paths, fault simulation, index statistics, and system index calculations. However, the calculation accuracy of the simulation method is inversely proportional to the calculation efficiency.

Therefore, some researchers focused on improving simulation method to reduce simulation variance and improve simulation efficiency. To the best of our knowledge, there are mainly stratified sampling method, control variable method, important sampling method, dual variable method, Latin hypercube sampling (LHS) method, and so on [38]. Zhang et al. [39] proposed an improved LHS method (hereinafter, ILHS) for evaluating the reliability of power system; this method combines LHS method and important method. In addition, some studies used the mixed method (hereinafter, MIXD) to evaluate the reliability of system, which combines the analytical method and Monte Carlo method. Han et al. [40] used analytical method to calculate

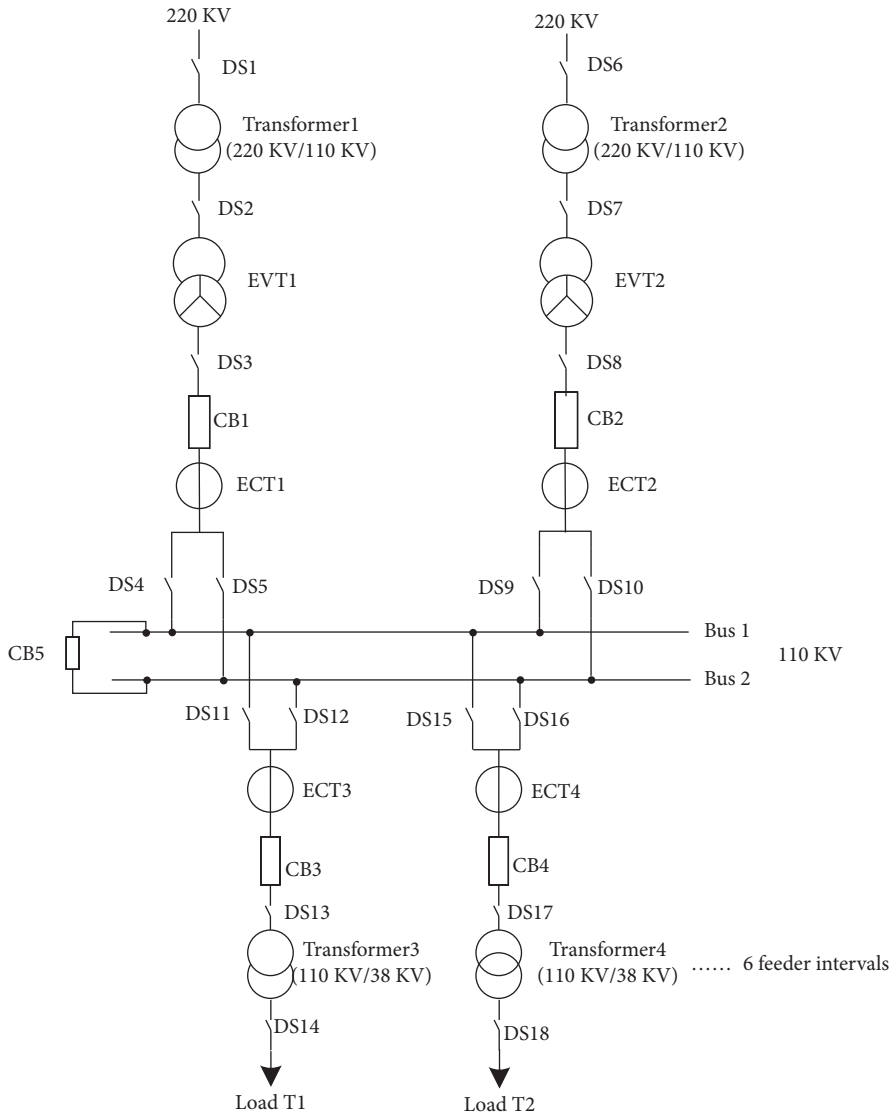


Figure 5: The main electrical connection of D2-1 smart substation.

the availability of protection and monitoring functions and then used the nonsequential Monte Carlo method for reliability evaluation. Hou et al. [41] used the mixed method to obtain a system state distribution, which is more close to the true distribution probability. Based on the impact analysis of failure mode. These methods improved the efficiency of reliability evaluation in different degrees.

In conclusion, the analytical method improves the simulation efficiency of Monte Carlo method by reducing the time of each state evaluation. And it does not affect the application of the variance reduction techniques. Thus, the analytical method can be combined with the variance reduction methods, which not only shortens the time of each state evaluation, but also reduces the simulation variance.

5.2.2. MALI-Hybrid. According to the analysis in Section 5.2.1, we present MALI-hybrid method, which combines the Monte Carlo method, analytical method, LHS method, and important sampling method. This method improves the simulation efficiency in two aspects, which not only shortens the time of each state evaluation, but also reduces the simulation variance and improves simulation convergence speed.

The procedure of MALI-hybrid is shown in Figure 7, in which f_{sampling} is the function of the improved LHS method, and its procedure is shown in Figure 8. We implemented MALI-hybrid in MATLAB to verify these two procedures.

In Figure 7, "Generate the minimum cut sets of the system" refers to getting the minimum cut sets of the smart substation based on Figure 6 and the minimum cut set theory of analytic method. The generated sets are {L1, L2}, {L1, L4}, {L2, L3}, {L3, L4}, {L5, L11}. {L5}, {L6}, {L6, L11}, {L7, L8}, {L7, L9}, {L8, L9}, and {L9, L10}. The different sets are connected in series. If any set fails, the system fails. Based on adjacency matrix method, we have programmed to generate the minimum cut sets in MATLAB. Thus, the obtained minimum cut sets are connected as shown in Figure 9.

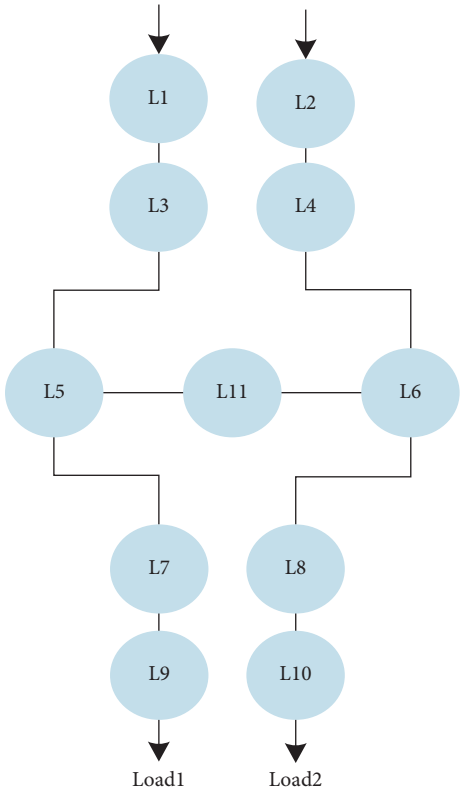


Figure 6: Logic diagram of the main electrical connection.

In Figure 8, "Parameter initialization" refers to the set initial multiplier K_0 as 1.1 by referring to [39], the important sampling times Num as 5000, and the maximum sampling times $N_{\max}$ as 10000. K is the optimal multiplier, which can be calculated by equations (14) and (15). $|K - K_0| \le 0.01$ refers to the criteria of optimizing K. The criteria are set as 0.01 by referring to [39]. The criteria value (i.e., 0.01) refers to the calculating accuracy. The smaller the criteria value is, the higher the calculating accuracy is, and the more simulation time is needed. N_0 are the iteration times.

$$K = -\frac{B + \sqrt{B^2 - AC}}{A}, \tag{14}$$

$$\begin{cases} A = \dfrac{n_1}{n_0 + n_1}\overline{\lambda} - \left(1 - \dfrac{n_1}{n_0 + n_1}\right)\overline{\lambda}(1 - \overline{\lambda}), \\ B = -\dfrac{n_1}{n_0 + n_1}\overline{\lambda}, \\ C = \dfrac{n_1}{n_0 + n_1}, \\ \overline{\lambda} = \dfrac{1}{n}\sum_{i=1}^{n}\lambda_i, \end{cases} \tag{15}$$

where n is the total number of elements in the system; n_0 and n_1 are the numbers of normal elements and faulty elements, respectively; $\overline{\lambda}$ is the average failure rate of all elements.

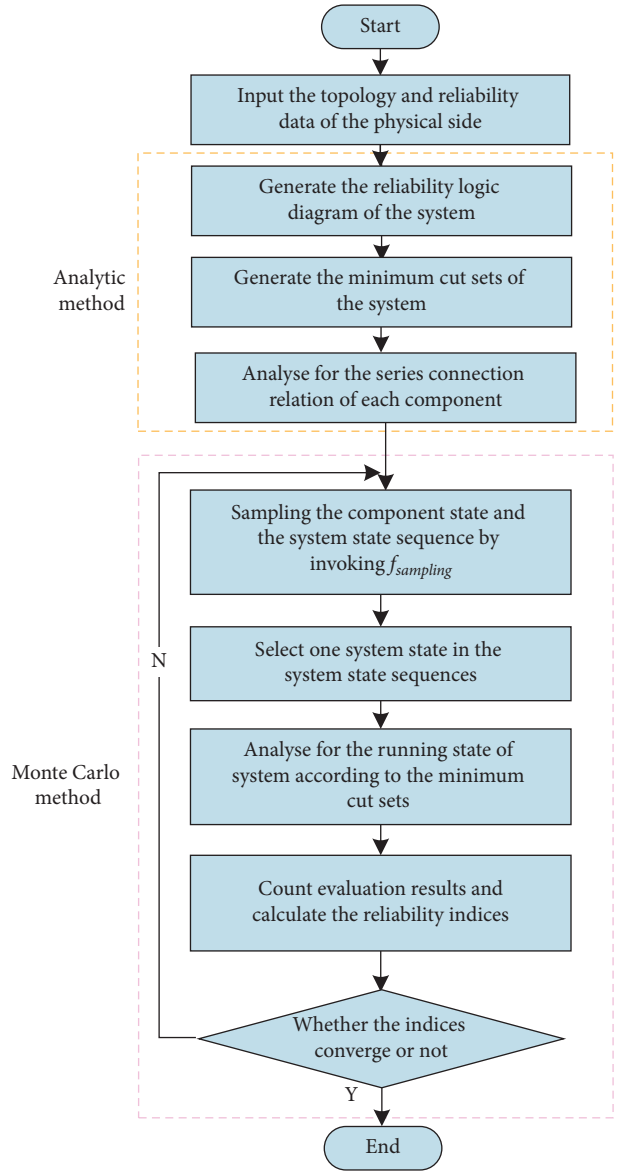


Figure 7: Procedure of the presented MALI-hybrid method.

6. Case Study

The case study in this section is based on the logic correlation diagram of the main electrical connection shown in Figure 6.

6.1. Reliability Calculation of Elements in Smart Substation. Based on the data about failure rate and repair rate in references [42–44], the reliabilities of physical equipment calculated by equation (5) are shown in Table 2. In actual evaluation, the connecting wires between physical equipment in smart substation are very short. Therefore, the reliability of electrical line is set as 1 in the following calculations, which is not presented in the table. Similarly, the reliabilities of cyber equipment are shown in Table 3.

We apply AHP model shown in Figure 3 to calculate the importance degree ∂_i. The results are shown in Table 4.

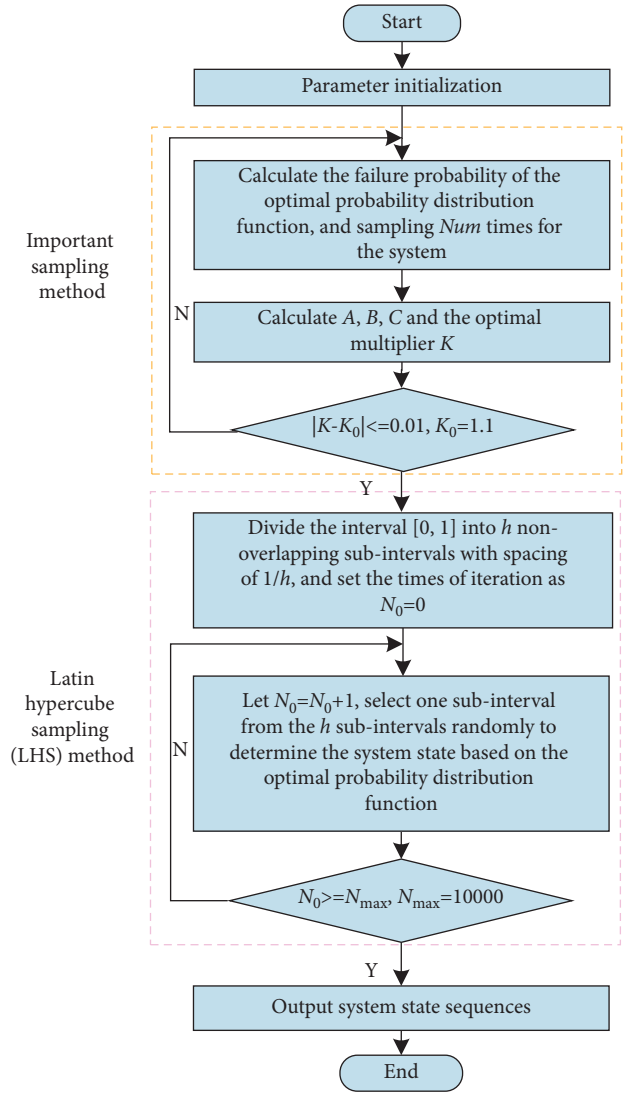


Figure 8: Procedure of the improved LHS method.

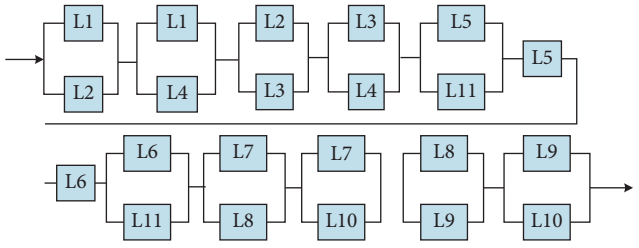


Figure 9: Connection between the minimum cut sets.

Table 2: Reliability data of physical elements.

Device name	Failure rate (a^{-1})	Repair rate (a^{-1})	Reliability
Bus	0.1	2920	0.999966
DS	0.015	584	0.999974
CB	0.12	219	0.999452
Transformer	0.03	97.3	0.999692
ET	0.014	876	0.999984

Table 3: Reliability data of cyber elements.

Device name	Failure rate (a^{-1})	Repair rate (a^{-1})	Reliability
Switch	0.02	365	0.999945
P&C IED	0.0067	365	0.999982
MU	0.0067	365	0.999982
CB IED	0.005	365	0.999986
Communication link	0.001	365	0.999997

Table 4: Importance degree of the cyber equipment.

Device name	Importance
Switch	0.2476
P&C IED	0.3125
MU IED	0.1768
CB IED	0.2631

Table 5: Equivalent reliability of cyber equipment considering importance degree or packet error rate.

Device name	Reliability
Switch	0.999986
P&C IED	0.999994
MU IED	0.999997
CB IED	0.999996
Communication link	0.999997

The reliability of data transmitting on communication links can be calculated by equation (8) based on packet error rate. Because the communications in smart substation is based on Ethernet, the packet error rate is equal to the omission rate of cyclic redundancy check (CRC) used in Ethernet (i.e., $P_e = 2E - 10$). Thus, the reliability of transmitted data is $D = 1 - (2E - 10)$. The equivalent reliabilities of cyber elements are calculated according to equations (5)–(8), and the results are shown in Table 5.

According to Section 4, the equivalent reliability of CB considering the reliability factors of the cyber side and cyber-physical interactions can be calculated by equations (9)–(12). The error rate of measurement information is set as 0.0001% in this calculation because of no relevant statistical data. The results are shown in Table 6.

6.2. Reliability Evaluation of Smart Substation When Different Factors Are Considered

6.2.1. Rationality Verification for Indices. To verify the rationality of the reliability indices proposed in this paper, the following three scenarios are presented.

Scenario 1: Assume that the cyber side and the measurement information of the physical side are completely reliable. Therefore, the reliability evaluation of smart substation is conducted only considering the factors of the physical side.

Table 6: Equivalent reliabilities of CBs considering the reliability factors of the cyber side and interactions.

Condition	Type of CB	Equivalent reliability
Reliability factors of the cyber side is considered	Bus CB	0.999342
	Feeder CB	0.999281
	Transformer CB1	0.999337
	Transformer CB2	0.999342
Interactions is considered	Bus CB	0.999325
	Feeder CB	0.999264
	Transformer CB1	0.999321
	Transformer CB2	0.999325

Table 7: Results of reliability evaluation based on the MALI-hybrid method in this paper.

Scenario	Reliability	Condition
Scenario 1	0.999896	Only the factors on the physical side are considered
Scenario 2	0.999626	The effect of the cyber side on the physical side is considered
Scenario 3	0.999326	The cyber-physical interactions are considered

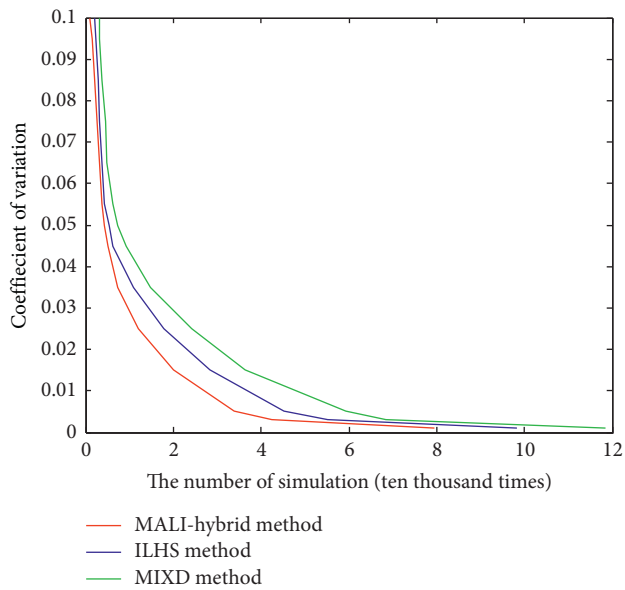


Figure 10: The dynamic variation diagram of variance coefficient with the number of simulations.

Scenario 2: Assume that the measurement information of the physical side is completely reliable, but the cyber side is not. Therefore, the reliability evaluation of smart substation is conducted considering the effect of the cyber side on the physical side.

Scenario 3: Assume that both the cyber side and the measurement information of the physical side are not reliable. Therefore, the reliability evaluation of smart substation is conducted considering the interactions between the cyber side and the physical side.

Table 7 shows the evaluation results of smart substation in the three scenarios based on the data in Section 6.1 and the MALI-hybrid method presented in Section 5.2.2.

Table 7 shows that the reliability of the smart substation in Scenario 1, Scenario 2, and Scenario 3 is successively reduced, which is consistent with the actual situation. Therefore, the presented reliability indices in this paper are rational and feasible.

6.2.2. Efficiency Verification for MALI-Hybrid. According to the analysis in Section 5, the MALI-hybrid method can reduce the simulation variance and improve simulation speed. Thus, to verify the efficiency of the MALI-hybrid method, we evaluated the reliability of the smart substation in Scenario 3 from Section 6.2.1 using the MALI-hybrid method presented in Section 5.2.2, ILHS method [33], and MIXD method [35], respectively. The dynamic variation of variance coefficient with the number of simulations is shown in Figure 10.

As shown in Figure 10, MALI-hybrid method has a smaller variance coefficient than the other two methods when the numbers of simulations are the same. It shows that

Table 8: Comparison of the MALI-hybrid method in Section 5.2.2 and Monte Carlo method.

Numbers of simulation	Reliability evaluation results based on the presented method	Reliability evaluation results based on the Monte Carlo method
2000	0.999989	0.999289
4000	0.999759	0.999269
6000	0.999659	0.999259
8000	0.999539	0.999239
10000	0.999326	0.999229

MALI-hybrid method does improve the simulation efficiency compared to the other two methods.

6.2.3. Correctness Verification for MALI-Hybrid. To verify the correctness of MALI-hybrid method, we evaluated the reliability of the smart substation in Scenario 3 from Section 6.2.1 using the MALI-hybrid method and Monte Carlo method. The results are shown in Table 8.

By analyzing the results from Table 8, it can be seen that the reliability difference between the MALI-hybrid method and Monte Carlo method is small and within the engineering error range. That is, the calculation precision of the MALI-hybrid method is reasonable. Therefore, MALI-hybrid method presented in Section 5.2.2 is correct.

7. Conclusions

This paper studies the reliability evaluation method for smart substation from the perspective of cyber-physical system. We point out that reliability evaluation of smart substation should consider the reliability factors on both the cyber side and the physical side, and their interactions. And the problem how cyber-physical interactions affect the reliability of smart substation can be converted into analyzing the effect of measurement and control messages on CBs' action. Based on this view, new reliability indices for smart substation are proposed: (1) reliability index of cyber equipment considering its effect on the communication business when the data leakage, packet error, message tampering or forgery, packet loss or interruption, and message delay occur, (2) reliability index of communication links considering the reliability of the transmitted data on them, and (3) equivalent reliability indices of CB considering the cyber-physical interactions. Furthermore, a MALI-hybrid reliability evaluation method is presented. This method combines the Monte Carlo method, analytical method, LHS method, and important sampling method. The case studies show that the indices and the method in this paper are reasonable, and our MALI-hybrid method is superior to the existing methods in terms of simulation speed.

In the future, we will further conduct a study on smart grid from the perspective of cyber-physical system based on this work, so as to propose broadly applicable reliability evaluation indicators and methods for smart grid.

Data Availability

The data used to support the findings of this study are currently under embargo, while the research findings are commercialized. Requests for data after publication of this article will be considered by the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Research on the evolution mechanism and early defense of cascading failures across spaces caused by coordinated cyberattacks in Grid CPS, No. 51977155).

References

[1] A. Zheng, Q. Huang, and D. Cai, "Quantitative assessment of stochastic property of network-induced time delay in smart substation cyber communications," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2407–2416, 2020.

[2] L. J. Chen, S. W. Mei, and Y. Chen, "Smart grid information security and its influence on power system survivability," *Control Theory & Application*, vol. 29, no. 2, pp. 240–244, 2012.

[3] Y. Tang, Q. Chen, and M. Y. Li, "The research overview of network attack in Cyber-Physical System," *Automation of Electric Power Systems*, vol. 40, no. 17, pp. 59–69, 2016.

[4] Y. M. Xia, *Performance analysis of communication networks and cyber security assessment of power systems*, Ph.D. Thesis, ZheJiang University, Zhejiang, China, 2015.

[5] Y. Zhou and Z. H. Chen, "The analysis of 7.10/central China (Henan) grid accident," *Hunan Electric Power*, vol. 28, no. 3, pp. 28–31, 2008.

[6] Q. L. Guo, S. J. Xin, J. H. Wang, and H. B. Sun, "Comprehensive security assessment for cyber physical energy system: a Lesson from Ukraine's blackout," *Automation of Electric Power Systems*, vol. 40, no. 5, pp. 145–147, 2016.

[7] Z. H. Li, *Research on the key techniques of performance evaluation system for electronic transformers*, Ph.D. Thesis, Huazhong University of Science and Technology, Wuhan, China, 2014.

[8] Y. Cai, Y. Chen, Y. Li, and Y. J. Cao, "Reliability analysis of cyber–physical systems: case of the substation based on the IEC 61850 standard in China," *Energies*, vol. 11, no. 10, p. 2589, 2018.

[9] Y. L. Zhang, "*The reliability and related issues of secondary system of intelligent substation*," M.S. Thesis, Shanghai Jiao tong University, Shanghai, China, 2013.

[10] Y. Tang, X. Han, and Y. J. Wu, "Electric power system vulnerability assessment considering the influence of communication system," *Proceedings of the CSEE*, vol. 35, no. 2, pp. 6066–6074, 2015.

[11] H. T. Lei, C. Singh, and S. Aprintson, "Reliability modeling and analysis of IEC 61850 based substation protection systems," *IEEE Transactions on Smart Grid*, vol. 5, no. 5, pp. 6066–6074, 2014.

[12] C. W. Ten, K. Yamashita, Z. Y. Yang, A. V. Vasilakos, and A. Ginter, "Impact assessment of hypothesized cyberattacks on interconnected bulk power systems," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4405–4425, 2018.

[13] W. X. Liu, Q. Gong, H. Han, Z. Q. Wang, and L. F. Wang, "Reliability modeling and evaluation of active cyber physical distribution system," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7096–7108, 2018.

[14] V. Aravinthan, T. Balachandran, and M. Ben, "Reliability modeling considerations for emerging cyber–physical power systems," in *Proceedings of the 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pp. 24–28, Boise, ID, USA, June 2018.

[15] R. Wang, Q. Y. Sun, and W. Hu, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, pp. 7866–7879, 2021.

[16] Q. Y. Sun, R. K. Han, and H. G. Zhang, "A multiagent-based consensus algorithm for distributed coordinated control of distributed generators in the energy internet," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 3006–3019, 2015.

[17] R. Wang, Q. Y. Sun, and D. Z. Ma, "The small-signal stability analysis of the droop-controlled converter in electromagnetic timescale," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1459–1469, 2019.

[18] R. Wang, Q. Y. Sun, D. Z. Ma, and X. G. Hu, "Line impedance cooperative stability region identification method for grid-tied inverters under weak grids," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 2856–2866, 2020.

[19] W. Hu, C. H. Ruan, and H. Nian, "Simplified Modulation scheme for open-end winding PMSM system with common dc bus under open-phase fault based on circulating current suppression," *IEEE Transactions on Power Electronics*, vol. 35, no. 1, pp. 10–14, 2020.

[20] X. G. Hu, H. G. Zhang, D. Z. Ma, and R. Wang, "Hierarchical pressure data recovery for pipeline network via generative adversarial networks," *IEEE Transactions on Automation Science and Engineering*, 2021, inpress.

[21] D. Z. Ma, Y. B. Li, X. G. Hu, H. Zhang, and X. Xie, "An optimal three-dimensional drone layout method for maximum signal coverage and minimum interference in complex pipeline networks," *IEEE Transactions on Cybernetics*, 2021, inpress.

[22] Y. G. Chen, J. E. Li, and A. Xu, "Reliability evaluation method for smart substation based on cyber-physical fusion," *Automation of Electric Power Systems*, vol. 45, no. 3, pp. 102–110, 2021.

[23] IEC, *Communication Network and Systems in Substations-Part 5: Communication Requirements for Function and Device Models*, IEC, Geneva, Switzerland, 2003.

[24] H. C. Zhao, *Analytic Hierarchy Process*, Science Press, Beijing, China, 1986.

[25] B. Falahati, Y. Fu, and L. Wu, "Reliability assessment of smart grid considering direct cyber-power interdependencies," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1515–1524, 2012.

[26] S. Liu, S. Mashayekh, D. Kundur, T. Zourntos, and K. Butler-Purry, "A framework for modeling cyber-physical switching attacks in smart grid," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 2, pp. 273–285, 2013.

[27] Y. C. Zhang, L. F. Wang, and W. Q. Sun, "Trust system design optimization in smart grid network infrastructure," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 184–195, 2013.

[28] K. R. Chen, F. S. Wen, and J. H. Zhao, "Vulnerability assessment of cyber-physical power system considering virtual cyber-physical connections," *Electric Power Automation Equipment*, vol. 37, no. 12, pp. 67–72, 2017.

[29] S. Wang and X. M. Huang, "Filtering methods of GOOSE messages," *Automation of Electric Power Systems*, vol. 32, no. 19, pp. 54–57, 2008.

[30] X. G. Yin, Z. Zhang, and H. Y. Wang, "Multiple transformers and CTs interactive effect and its impact on the protective relay," *Power System Protection and Control*, vol. 44, no. 23, pp. 1–9, 2016.

[31] Y. Zhao and K. G. Xie, "An analytical approach to compute the probability density distributions of reliability indices for bulk power systems," *Proceeding of the CSEE*, vol. 31, no. 4, pp. 31–38, 2011.

[32] Y. Liu, "*Technology and economy evaluation of the new generation intelligent substation electrical main wiring design scheme*," M.S. Thesis, North China Electric Power University, Beijing, China, 2017.

[33] W. X. Liu, Q. Gong, and J. Guo, "Reliability simulation of AND cyber-physical system based on hybrid communication network," *Proceedings of the CSEE*, vol. 36, no. 6, pp. 1706–1718, 2018.

[34] C. Wang, T. Zhang, and F. Luo, "Impacts of cyber system on microgrid operational reliability," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 105–115, 2019.

[35] R. D. Sun and K. G. Xie, "Reliability evaluation of distribution networks using Monte Carlo method considering correlations between wind speed and load," *Power System Protection and Control*, vol. 40, no. 18, pp. 12–18, 2012.

[36] J. R. Araújo, E. N. M. Silva, and A. B. Rodrigues, "Assessment of the impact of microgrid control strategies in the power distribution reliability indices," *Journal of Control, Automation and Electrical Systems*, vol. 28, no. 2, pp. 271–283, 2017.

[37] C. Zhou, X. Li, J. T. Wang, W. H. Liang, and W. Z. Yin, "Reliability calculation of distribution network based on sequential Monte Carlo simulation of load path," *Mechanical and Electrical Information*, vol. 12, pp. 24–27, 2019.

[38] X. Z. Xie, "Reliability improvement and its assessment method for distribution network," *Electrical Equipment*, vol. 9, pp. 76–78, 2004.

[39] W. F. Zhang, Y. B. Che, and Y. S. Liu, "Improved Latin hypercube sampling method in power system reliability evaluation," *Automation of Electric Power Systems*, vol. 39, no. 4, pp. 52–57, 2015.

[40] Y. Q. Han, Y. F. Wen, C. X. Guo, and H. Huang, "Incorporating cyber layer failures in composite power system reliability evaluations," *Energies*, vol. 8, no. 9, pp. 9064–9086, 2015.

[41] Y. S. Hou, X. L. Wang, and Y. Zhang, "Dimensional importance based quasi-Monte Carlo for power system reliability evaluation," *Automation of Electric Power Systems*, vol. 40, no. 16, pp. 31–37, 2016.

[42] P. Brant, *The Automation of Substations*, pp. 191–200, China Electric Power Press, Beijing, China, 2009.

[43] L. Cheng and J. He, *Power System Reliability Principle and Application*, pp. 356–358, Tsinghua University Press, Beijing, China, 2015.

[44] T. W. Wang, M. Xie, and Y. Q. Sun, "Analysis of reliability for relay protection systems in smart substation," *Power System Protection and Control*, vol. 43, no. 6, pp. 58–66, 2015.

Hindawi
Complexity
Volume 2021, Article ID 5578972, 9 pages
https://doi.org/10.1155/2021/5578972

Research Article

Vehicle-Mounted Photovoltaic System Energy Management in Intelligent Transportation Systems: A Maximum Power Point Tracking Control

Jingao Wang,[1] Qifei Liu ,[1] and Silan Jing[2]

[1]*College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China*
[2]*College of Business Administration, Northeastern University, Shenyang, Liaoning 110819, China*

Correspondence should be addressed to Qifei Liu; lqf184848458@163.com

Received 22 January 2021; Revised 26 May 2021; Accepted 24 July 2021; Published 5 August 2021

Academic Editor: Dan Selisteanu

Electric vehicles have become the main contributor in terms of reducing fuel consumption and CO_2 emission. Although the government is vigorously promoting the use of electric vehicles worldwide, the range anxiety still impedes the rapid development of electric vehicles, especially when air-conditioning also adds battery power consumption and aggravates the range anxiety. To this end, this paper proposes an improved vehicle-mounted photovoltaic system energy management in intelligent transportation systems, which is a maximum power point tracking control system. Meanwhile, since the power of solar panels is usually relatively small and the power changes at any time, low power density and poor controllability are difficult to avoid. In order to solve this problem, this paper offers a tracking control method to improve the output efficiency of solar panels. For improving photovoltaic conversion efficiency and maximizing output power, traditional photovoltaic power panels are often dominated by a centralized maximum power point tracing control, which is named MPPT. Although the cost under this case is lower, the output power of all photovoltaic panels cannot be maximized under the condition of uneven illumination or local mismatch. To improve the situation, a micro-scale inverter is proposed to provide MPPT control of photovoltaic modules, which can effectively improve the output power of each photovoltaic panel. Moreover, our MPPT algorithm is applicable to cloud shadow, building shadow, and shade, and it is more suitable for the car roof. Firstly, the Diode 5-parameter model is used to deduce the *I-U* equation of the photovoltaic module considering shadow shading, and then the real-time 5 parameter equation is formed by using the measured data group and selected. The reasonable initial value is used to iteratively solve the real-time value of 5 parameters, which is further to judge the masking situation. The maximum power point (MPP) is solved directly by the mathematical method based on the mathematical model of *I-U* relation mathematics, and the DC-DC circuit is used to adjust the running point to MPP. Unlike the traditional MPPT method, the method in this paper is based on the physical model of solar cells, and MPP tracking is based on mathematical methods. Based on this, it does not need to cause multiple interference to the circuit, and the tracking efficiency is high. Finally, the relative experimental results are provided to verify the performance of the proposed method.

1. Introduction

Because of the high temperature in a vehicle in summer, the vehicle owner is very uncomfortable when entering it. In particular, when children are locked in a car, they are frequently killed by heat. Based on this problem, using a photovoltaic cell to drive the air-conditioner and cool the car is very effective. The cooperative control of photovoltaic modules such as solar cells is the key to achieve it because photovoltaic panels can get extra energy and direct power to the air-conditioner without using the electricity in the dynamic battery, which can reduce mileage anxiety. However, the photovoltaic panels provide very limited power due to the limitation of the car roof area. Thus, it is significant to find the maximum power point of onboard photovoltaic panels to increase its power generation and apply this method to the solar energy controller in Figure 1. In this section, the most important difficulty to realize cooperative

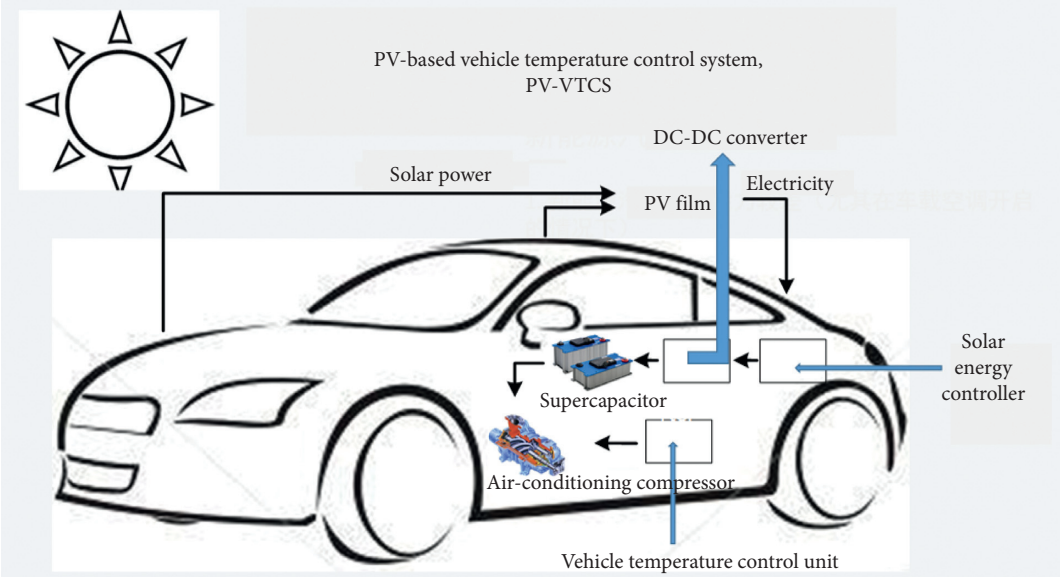


Figure 1: PV-based vehicle temperature control system.

control is how to track the maximum power point of photovoltaic panels.

According to the external environment, the output power of solar cell varies with the operating point of voltage and current, but the output power has a maximum value, and at this time, the photoelectric conversion efficiency reaches the peak [1]. Therefore, photovoltaic development should be specialized in maximum power point tracking control to raise the efficiency in solar power stations [2]. Since the solar cell is a nonlinear system, the relationship between the output current and voltage is a complex implicit nonlinear system [3], so it is difficult to use mathematical model tracking (MPPT), the maximum rate of photovoltaic directly [4].

The general MPPT method does not need to consider the intelligent disturbance optimization control of the model, such as the disturbance observation method increasing conductance method [5, 6]. The advantage is they do not consider the complexity of the model, and logic is simple [7], but the disadvantage is that it will cause multiple disturbances of the circuit, the control speed is slow, and the adjustment cannot be realized in one step [8]. The improved disturbance observation method can improve the tracking speed by a reasonable adaptive step size control strategy, such as changing the disturbance step size according to the slope of the *P-U* curve, and can alleviate the oscillation of the circuit at the maximum power point [9].

The traditional centralized inverter utilizes a master MPPT controller to control the total output power of the photovoltaic power station by disturbance observation [10]. At this point, the current constraint between the series solar panels in the photovoltaic array and the voltage constraint between the parallel groups will lead to this condition. While the total output power reaches the top, the corresponding output power of each solar panel does not [11] at its respective maximum power points, especially when the illumination and cloud shadow obscured, or other mismatch occurs [12]. Although some modern algorithms work very well in tracking the maximum power points [13–15], these methods can only be used for static photovoltaic films, and the equipment cost is high. And also, they are suitable for large-scale photovoltaic power generation but are not used for vehicles. As a consequence, a sort of inverter for controlling a series of photovoltaic panels and a microinverter for controlling solar cell veneer are proposed in this paper. Both deeply enhance the utilization rate of illumination; the former cost is lower, but the latter has higher reliability and photoelectric conversion rate. Expansion is easier too [16].

Aiming at the vehicle microinverter installed in the solar cell veneer, this paper considered its mechanism mathematical model and quantified the process of finding the maximum power point. The maximum power point of three solar panels is tracked by the mathematical method by using the measured data. This avoids the traditional MPPT method to the circuit caused by multiple disturbances, which is more suitable for improving the power generation efficiency of vehicle photovoltaic panels with less power.

2. Five-Parameter Model for Photovoltaic Modules

2.1. Five-Parameter Model for Photovoltaic Modules. Figure 2 shows the single diode equivalent circuit of the solar cell. It can be seen that the solar cell is equivalent to a current source parallel to a diode considering the series parallel loss. In Figure 2, I, I_{ph}, I_D, and I_{sh} stand for the output current, the photogenerated current, the equivalent diode current, and the current flowing over the equivalent parallel resistor, respectively. R_{sh}, R_s are equal to parallel resistance and series resistance; U_D, U are equal to diode voltage and output voltage, respectively.

The equivalent circuit corresponds to five parameter model expressions:

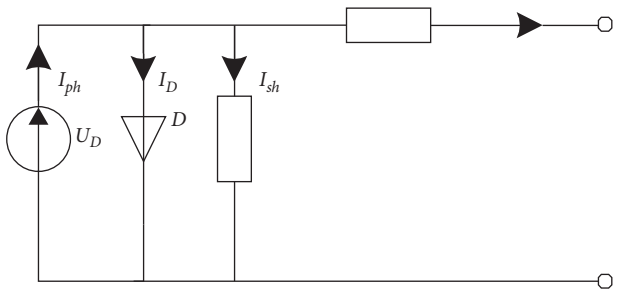


Figure 2: Equivalent circuit of the photovoltaic cell.

$$I = I_{ph} - I_s\left[e^{\left(q\left(U+IR_S\right)/AkT\right)} - 1\right] - \frac{U + IR_S}{R_{sh}},$$
$$I = I_{ph} - I_s\left[e^{q\left(U+IR_S\right)/AkT} - 1\right] - \frac{U + IR_S}{R_{sh}}. \quad (1)$$

In this form, q is the electrical power constant, whose value is 1.602×10^{-19} C; k is the Boltzmann constant, whose value is 1.381×10^{-23} J/K; A is the diode characteristic fitting coefficient, whose value is from 1 to 2; and I_s is the equivalent diode saturated current.

Formula (1) can be regarded as an equation with five undetermined parameters, which are I_{ph}, R_{sh}, a, I_s, and R_s (a represents comprehensive fitting coefficient, equal to q/AkT and T represents real-time temperatures)

2.2. Relation of Parameters to Irradiance and Temperature. The correction at different temperatures and irradiance form is

$$I_{ph} = \left(\frac{S}{S_{\text{ref}}}\right)\left[I_{ph,\text{ref}} + C_T\left(T - T_{\text{ref}}\right)\right],$$
$$I_s = I_{s,\text{ref}}\left(\frac{T}{T_{\text{ref}}}\right)e^{\left[\left(qE_g/kA\right)\left(\left(1/T_{\text{ref}}\right)-\left(1/T\right)\right)\right]},$$
$$R_S = R_{s,\text{ref}}\left(\frac{T}{T_{\text{ref}}}\right)^3\left[1 - \beta\ln\left(\frac{S}{S_{\text{ref}}}\right)\right], \quad (2)$$
$$R_{sh} = R_{sh,\text{ref}}\left(\frac{S}{S_{\text{ref}}}\right).$$

In this form, $I_{ph,re}$ is photogenerated current under standard operating conditions; S equals real-time irradiance; S_{ref} is irradiance under standard conditions ; C_T is temperature coefficient, available from the manufacturer; I_{ref} is diode saturation current in standard condition; T_{ref} is temperature under standard conditions; and E_g is the bandgap, depending on the solar cell material. In this paper, the bandgap width of silicon is 1.12 eV; β is the correction factor; and β is 0.217.

The variation of these five parameters with temperature and irradiance is investigated, and the parameter values under standard conditions (ring temperature 25°C, illumination 1000 W/m^2) are compared and calculated:

(1) Temperature is 0.99 times of standard temperature and irradiance

(2) Irradiance is the standard value and the temperature is 1.01 times the standard condition (the temperature change is about 3 K, considering the temperature difference within 1 min in extreme cases)

(3) The temperature is the standard temperature, the irradiance is the parameter value of 0.7 times under the standard condition (the irradiance variation is 300 W/m^2, considering the amplitude difference within 1 h under normal conditions)

(4) The irradiance is the standard value and the temperature is 1.03 times the parameter value under the standard condition (the temperature change is about 9 K, considering the temperature difference within 1 h in extreme cases)

From Table 1, it can be seen that when the time scale is measured in minutes, the variation of each parameter with irradiance and temperature can be ignored. If the time scale is counted in hours, the change rate of each parameter is much smaller than that of irradiance, so the effect of temperature can be ignored within 1 h. Meanwhile, in the 5 parameters including a, I_s, and R_s, the irradiance has little effect, and it can be considered as a definite value; it means that only I_{ph} and R_{sh} changed within 1 h. Two calculation modes can be divided according to the time scale. If considering the whole point of timing (such as 10 : 00~11 : 00), the time length between every two whole points can be specified as a large time scale; then, the large time scale is divided into 60 parts. That is 1 h per minute.

Within the same large time scale, each small time scale corresponds to different I_{ph}, R_{sh}. At this time, only the I_{ph}, R_{sh} parameters should be identified to simplify the calculation and improve the calculation speed. The whole point needs to consider the large time scale and a, I_s, and R_s change accordingly. At this time, we need to calculate 5 parameters.

2.3. Collaborative 5-Parameter Model of N Photovoltaic Modules. The photovoltaic module containing N monomers is demonstrated in Figure 3(a). Each monomer is divided into M groups, each of which is connected with a bypass diode. An equivalent circuit of a series unit on the component board is shown in Figure 3(b).

Considering the masking situation, when one or several monomers in a set are obscured, the current will go through directly through the bypass diode. The terminal voltage of each group is the forward conduction voltage of the diode, and the silicon tube is 0.3 V, so the conduction voltage $U_{DD} = 0.3$ V.

Taking into account $N = 60$, $M = 20$ groups of panels, a solar panel has N_1 cells without shelter, and N_2 cells with shelter; the equivalent diode voltage of each group of cells without shelter is shown in the following formula:

Table 1: Relationship between the parameters and conditions.

Number of different environmental conditions	Temperature irradiance conditions	I_{ph} (A)	I_s (mA)	a	R_s (mW)	R_{sh} (W)
Standard situation	$S = S_{ref}$, $T = T_{ref}$	$I_{ph,ref}$	$I_{s,ref}$	a	$R_{s,ref}$	$R_{sh,ref}$
1	$S = 0.99S_{ref}$, $T = T_{ref}$	$0.99I_{ph,ref}$	$I_{s,ref}$	a	$1.002R_{s,ref}$	$0.99R_{sh,ref}$
2	$S = S_{ref}$, $T = 1.01T_{ref}$	$I_{ph,ref} + 0.01C_T$	$1.03I_{s,ref}$	$0.99a$	$1.01R_{s,ref}$	$R_{sh,ref}$
3	$S = 0.7S_{ref}$, $T = T_{ref}$	$0.7I_{ph,ref}$	$I_{s,ref}$	a	$1.08R_{s,ref}$	$0.7R_{sh,ref}$
4	$S = S_{ref}$, $T = 1.03T_{ref}$	$I_{ph,ref} + 0.03C_T$	$1.09I_{s,ref}$	$0.97a$	$1.03R_{s,ref}$	$R_{sh,ref}$

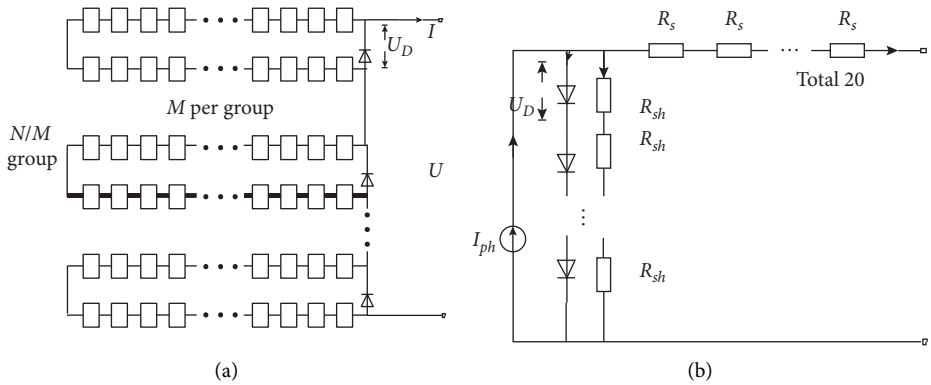


Figure 3: Schematic diagram of PV panel and equivalent circuit of a unit in PV pane. (a) Single board diagram. (b) Schematic diagram of the equivalent circuit of a series unit of solar cell veneer and component board.

$$U_D = \frac{U - (N_2/20)U_{DD} + IN_1R_S}{N_1} = \frac{U}{N_1}\frac{N_2U_{DD}}{20N_1} + IR_S. \tag{3}$$

Considering the U-I relation in the case of shelter,

$$I = I_{ph} - I_S\left[e^{a\left((U/N_1)-\left(N_2U_{DD}/20N_1\right)+IR_S\right)} - 1\right]\left(\frac{U}{N_1} - \frac{N_2U_{DD}}{20N_1} + IR_S\right)\frac{1}{R_{sh}}. \tag{4}$$

There are four different combinations for panels N_1 and N_2:

(1) $N_1 = 0$, $N_2 = 60$, $N_1/N_2 = 0$

(2) $N_1 = 20$, $N_2 = 40$, $N_1/N_2 = 12$

(3) $N_1 = 40$, $N_2 = 20$, $N_1/N_2 = 2$

(4) $N_1 = 60$, $N_2 = 0$, $N_1/N_2 = \infty$

When the battery chip is covered, if the U is normal, $N_1 = 60$, $N_2 = 0$; if the value of the U changes to about a third, $N_1 = 40$, $N_2 = 20$; if the U value changes to about 2/3 of the original, $N_1 = 20$, $N_2 = 40$; and if the U is very small, the panels do not work, considering the masking condition MPPT is not required. A five-parameter I-U relation can be obtained.

3. Analysis of the Characteristics of *P-U* Curves considering Shadows

3.1. Solar Cell P-U and I-U Analysis. P-U and I-U curves are determined under certain illumination and temperature conditions, as shown in Figure 4. PMP represents the power at the maximum power point, IMP represents the current at the maximum power point, and U_{MP} and U_{OC} represent the voltage at the maximum power point and the open-circuit voltage, respectively. There are 3 special points on the curve in Figure 4. Output short-circuit point corresponds to when the output voltage U is 0; output open-circuit point corresponds to when the output current is 0; and maximum power point corresponds to output power maximum point.

The power expression is shown in the following formula:

$$P = UI. \tag{5}$$

The derivative of the P to the U at the maximum power point is 0:

$$\frac{dP}{dU} = 0. \tag{6}$$

At this point, the output power is the maximum output power:

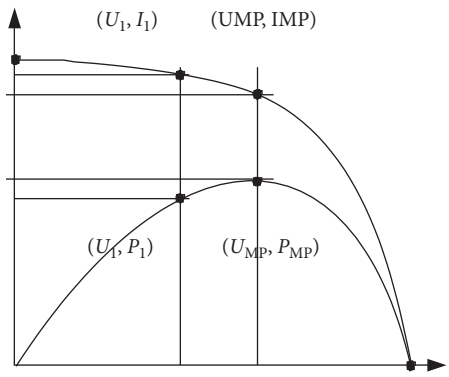


Figure 4: *I-U* curve and *P-U* curve of PV array.

$$P_{MP} = U_{MP} I_{MP}. \tag{7}$$

The photoelectric conversion rate and the utilization rate of light energy are the highest. When the initial operating point is (U_1, I_1) and the corresponding power is P_1, it can be seen from the *P-U* curve that the operating point can be adjusted to the maximum power point (U_{MP}, I_{MP}) through the external control device.

3.2. Bimodal Characteristics of Component Plate under Partial Shadow. When the solar panel is partially shaded, the *P-U* curve presents a bimodal characteristic, as shown in Figure 5, in which U_{JMP} represents the voltage at the local maximum power point and P_{JMP} represents the power at the local maximum power point. Conventional perturbation optimization M_{PPT} methods have the probability of tracing to the local peak points, which are the pseudo-maximum points.

4. Acquisition of Maximum Power Points

4.1. Parameter Identification. Aiming at the fault of parameter identification by parameter correction method, a parameter identification method which uses real-time gaugement data is proposed in this paper. According to the analysis in Section 3.2, the parameter identification is divided into two cases to simplify the computation:

(1) Adjust the external circuit of the photovoltaic module five times and then measure voltage and current values of 5 groups; thus, determine the corresponding 5-parameter model according to the voltage situation;

(2) According to the 5-parameter model and the 5 sets of voltage and current values, the 5 equations are determined as shown in formula (8). We take this approach to solve the 5 parameters and get the *U-I* relationship at the whole point. Iterative calculations only take advantage of the previous result at the whole point and then use it as the initial value

$$I_i = I_{ph} - I_S\left[e^{a\left((U/N_1)-\left(N_2U_{DD}/20N_1\right)+IR_S\right)} - 1\right] - \left(\frac{U_i}{N_1} - \frac{N_2U_{DD}}{20N_1} + I_iR_S\right)\frac{1}{R_{sh}}, \quad i = 1, 2, 3, 4, 5. \tag{8}$$

In this formation, I_i and I_s are the current values in the i set data, A is the saturation current of the equivalent diode; and U_i is the voltage value in group i data.

For any moment, assuming that the irradiance is the same as the previous integral moment, we replace 3 parameters of the five which are only affected by temperature change with integral values. By measuring two sets of voltage and current values, we established two equations for the two parameters greatly affected by irradiance. The expression of the two parameters is shown in formula (9), and we found the *U-I* relation at any time is obtained:

$$\begin{cases} I_{ph} = \dfrac{\left[I_1 + I_s\left(e^{V_1 - 1}\right)\right]\left(U_2 - \left(N_2U_{DD}/20\right) + N_1I_2R_S\right)}{\left(U_2 + N_1I_2R_S\right) - \left(U_1 + N_1I_1R_S\right)} - \dfrac{\left[I_2 + I_S\left(e^{V_2} - 1\right)\right]\left(U_1 - \left(N_2U_{DD}/20\right) + N_1I_1R_S\right)}{\left(U_2 + N_1I_2R_S\right) - \left(U_1 + N_1I_1R_S\right)}, \\ \dfrac{1}{R_{sh}} = \dfrac{N_1\left(I_1 - I_2\right) + N_1I_S\left(e^{V_1} - e^{V_2}\right)}{\left(U_2 + N_1I_2R_S\right) - \left(U_1 + N_1I_1R_S\right)}. \end{cases} \tag{9}$$

In this formula,

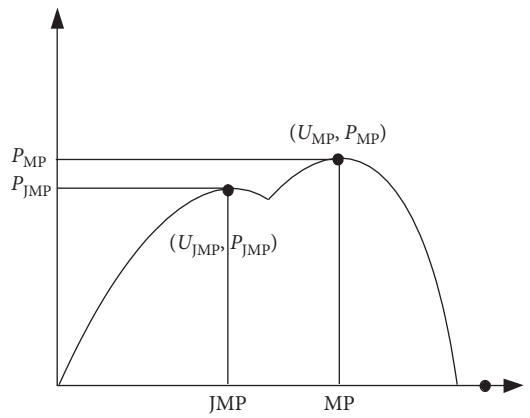


Figure 5: Bimodal characteristics of *P-U* curves of PV modules under shadow shading.

$$V_1 = a\left(\frac{U_1}{N} - \frac{N_1 U_{DD}}{20N_1} + I_1 R_S\right),$$
$$V_2 = a\left(\frac{U_2}{N} - \frac{N_1 U_{DD}}{20N_1} + I_2 R_S\right). \tag{10}$$

4.2. Maximum Power Point Positioning. Simplify the I-U curve formulas (4) to (11):

$$\ln\left[\frac{\left(I_{ph} - I\right)N_1 R_{sh} - \left(U - \left(N_2 U_{DD}/20\right) + IN_1 R_s\right)}{N_1 R_{sh} I_s} + 1\right] = a\left(\frac{U}{N_1} - \frac{N_2 U_{DD}}{20N_1} + IR_s\right). \tag{11}$$

Formulas (5) and (11) derive U on both sides separately and further simplify the following available formula:

$$\left[-\frac{U}{N_1} + I\left(R_s + R_{sh}\right)\right]\frac{1}{a} = \left(I_{ph} - I\right)R_{sh}\left(\frac{U}{N_1} - IR_s\right) - \left(\frac{U}{N_1} - \frac{N_2 U_{DD}}{20N_1} + IR_s + R_{sh} I_s\right)\left(\frac{U}{N_1} - IR_s\right). \tag{12}$$

Combine formulas (12) and (11), and the maximum power point (U_{MP}, I_{MP}) is solved out.

If assuming that

$$\frac{U}{N_1} - \frac{N_2 U_{DD}}{20N_1} + IR_s = x, \quad I = y, \tag{13}$$

then

$$\frac{U}{N_1} - IR_s = x + 2yR_s + \frac{N_2 U_{DD}}{20N_1}, \quad I = y,$$
$$U = \left(X + \frac{N_2 U_{DD}}{20N_1} - IR_s\right)N_1. \tag{14}$$

Formulas (11) and (12) can be reduced to the following formulas:

$$-\frac{1}{a}\left[x + y\left(2R_s + R_{sh}\right) + \frac{N_2 U_{DD}}{20N_1}\right] = \left[\left(I_{ph} - y\right)R_{sh} - x + R_{sh} I_s\right]\left(x + 2yR_s + \frac{N_2 U_{DD}}{20N_1}\right), \tag{15}$$

$$x(K+1) = \frac{AkT}{q}\ln\left\{\frac{\left[I_{ph} - y(K)\right]R_{sh} - x(K)}{R_{sh} I_s} + 1\right\}. \tag{16}$$

In these formulas, x and y are the replacement variables during the solution and K is the iteration number.

Formula (15) can be regarded as a quadratic equation about y, and it has 2 solutions generally, which proved the

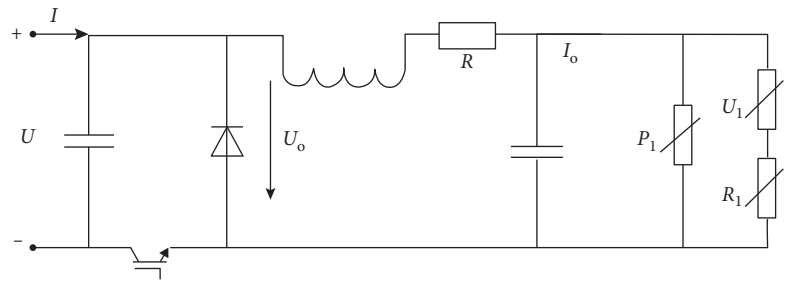


Figure 6: External buck circuit and load equivalent circuit diagram.

Table 2: Data of current and voltage of PV panel.

N_2/N_1	(U_1, I_1)	(U_2, I_2)	(U_3, I_3)	(U_4, I_4)	(U_5, I_5)
0	17.6, 5.3	23.0, 5.2	27.5, 5.0	29.1, 4.3	30.85, 3.6
0.5	11.4, 5.3	14.5, 5.2	17.9, 4.9	19.7, 4.2	21.0, 3.3
2	5.9, 5.2	7.11, 5.2	8.33, 5.0	9.22, 4.3	9.86, 4.0

bimodal characteristic mentioned in Section 3.2. It means that there are 2 points when the derivation of power to voltage is 0. Comparing the power value of the corresponding solution, we can abandon the "pseudo-maximum point."

Taking the maximum power point of a moment as the initial iteration value and putting $x(K)$, $y(K)$ to (16), we can find the corresponding $x(K_1)$ and then bring it back and find out $y(K_1)$. Under the condition that U and I are greater than 0, if $y(K_1)$ has two values, then compare the corresponding power values and take a set of values with the large corresponding power. Iterating it to the error within the allowable range, the corresponding I-U is the I-U at the maximum power.

4.3. Tracing the Maximum Power Point. Assuming that the load is a combination of constant power, invariant voltage, and invariant resistance, it is connected to the photovoltaic module by Buck circuit as shown in Figure 6.

The voltage and current at the end of the system in relation to the duty cycle α can be obtained from the equal power characteristic before and after the Buck circuit, as shown in the following formula:

$$\alpha = \frac{U_1 + \sqrt{U_1^2 + 4(UI + P_1)R_1}}{2U}. \tag{17}$$

After calculating the corresponding I-U value of the maximum power point, the corresponding duty cycle is procured, and the adjustment of the duty cycle can jump to the working point at the vicinity of the maximum power point directly.

5. Simulation Verification

Data in this article is simulated by PVsyst6.61. We establish a system that adopts a microinverter structure in PVsyst6.61 and each photovoltaic panel is equipped with a microinverter. Each PV module consists of 60 batteries.

We simulated a module board in PVsyst6.61 which has 6 series battery cells, and every two parallels have bypass diodes.

Select a panel in PVsyst to verify the feasibility of this method under the condition of no shading, 2 shadings, and 4 shadings, respectively. The simulation of cloud shadow masking and the internal mismatch is carried out through artificial masking, which causes some interference to the circuit. The data of panel current and voltage (3 groups in total) is collected as shown in Table 2.

By using the data in Table 2 and the quasi-Newton iteration method, the initial values of photovoltaic 5 parameters are $[I_{ph}, I_s, a, R_s, R_{sh}] = [6, 0.000001, 25, 0.0002, 6000]$. The initial values are gained from the factory parameters of photovoltaic modules, which is the corresponding 5 parameter values under the standard working condition. When the quasi-Newton method is used for iterative calculation, the selection of the initial values is related to the iterative speed and the convergence of the results. In the actual operation, the result of the last parameter identification can be taken as the initial value of the next iteration to ensure the reliability of the iteration and obtain the parameters. The identification results are shown in Table 3.

The photovoltaic characteristic curve equation can be written by using the 5 parameters. According to the steps described in Section 4.2, the maximum power points can be calculated by mathematical equations as shown in the MPP1 in Table 4. However, when the interference exists for a long time, the maximum power point tracked by the inverter actually can be measured as shown in the MPP2 in Table 4. Meanwhile, we also experimented this method under the condition of sunny days, cloudy days, and cloud shading days; the detailed data are in Table 5.

The value of the maximum power point calculated by the method described in this paper is consistent with the value of the maximum power point tracked by the MPPT control strategy basically. However, because there are errors during measure and model, the results will not be completely consistent.

Table 3: Result of parameter identification.

Experimental group	I_{ph} (A)	I_s (mA)	a	R_s (mW)	R_{sh} (W)
No shelter	5.4	2.2	25.5	3.3	12354
2-line shelter	5.3	2.1	25.5	4.2	8891
4-line shelter	5.2	2.5	25.7	3.8	9192

Table 4: Maximum power point computation.

Experimental group	I_{ph} (A)	U (V)	I (A)	P (W)	Relative power error (%)
No shelter	MPP1	28.2	4.8	135.36	−2.04
	MPP2	29.4	4.7	138.18	
2-line shelter	MPP1	18.2	4.9	89.18	7.39
	MPP2	17.3	4.8	83.04	
4-line shelter	MPP1	9.1	4.8	43.68	−2.1
	MPP2	9.5	4.7	44.65	

Table 5: Maximum power point computation under three different real-weather environments.

Experimental group	U (V)	I (A)	P (W)
Sunny days	27.92	4.65	129.83
Cloudy days	9.12	4.70	42.86
Cloud shading days	16.56	4.77	78.99

6. Conclusion

Using the five-parameter model of the solar cell to model the mechanism of photovoltaic module and the mathematical model to solve the maximum power point directly, the conclusion is as follows:

Different combinations of temperature and irradiance correspond to different values of 5 parameters; the disturbance circuit can gain equations for solving 5 parameters by measuring the values of voltage and current. In order to get the real-time value of 5 parameters, the maximum power point is solved directly by a mathematical equation. Compared with the traditional disturbance observation method, this direct mathematical method has fewer interference times with the circuit and it can provide faster MPPT speed.

The selection of initial values is related to the iteration speed and convergence of the results. The identification results of the parameters at the last moment offer the best iterative initial value for the next moment, and the iterative results have preferable convergence.

By dividing the unused time scale and utilizing the relationship between 5 parameters and temperature and irradiance, the parameters with less influence can be regarded as invariant for a period of time, so only two parameters need to be identified. Different time scales correspond to different parameter identification numbers; thus, this method simplifies the calculation process.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The project was supported by the Key Program of National Natural Science Foundation of China (61433004).

References

[1] M. A. Ghasemi, H. M. Foroushani, and F. Blaabjerg, "Marginal power-based maximum power point tracking control of photovoltaic system under partially shaded condition," *IEEE Transactions on Power Electronics*, vol. 35, no. 6, 2020.

[2] K. Liu and T. Z. Ji, X. Y. Li, X. F. He, and M. Zhang, Study of maximum power point tracking control based on GRNN neural network," *Advanced Materials Research*, vol. 989–994, pp. 3427–3432, 2014.

[3] F. Xu, R. Zhang, L. Wu, and H. Xu, "Self- adaption BP natural network for the maximum- power-point tracking in photovoltaic," *Acta Energiae Solaris sinica*, vol. 33, no. 3, pp. 468–472, 2012.

[4] J. Y. Shi, D. Y. Zhang, L. T. Ling et al., "Dual-algorithm maximum power point tracking control method for photovoltaic systems based on grey wolf optimization and golden-section optimization," *Journal of Power Electronics*, vol. 18, no. 3, pp. 841–852, 2018.

[5] S. K. Kollimalla and M. K. Mishra, "Variable perturbation size adaptive P&O MPPT algorithm for sudden changes in irradiance," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 3, pp. 718–728, 2014.

[6] S. Tang, Y. Sun, and Y. Chen, "An enhanced MPPT method combining fractional-order and fuzzy logic control," *IEEE Journal of Photovoltaics*, vol. 7, no. 2, 2017.

[7] S. VeerapenaH. Wena et al., "A novel global maximum power point tracking algorithm for photovoltaic system with variable perturbation frequency and zero oscillation," *Solar Energy*, vol. 181, pp. 345–356, 2019.

[8] S. A. Ansari and J. S. Moghani, "Soft switching flyback inverter for photovoltaic AC module applications," *IET Power Electronics*, vol. 13, p. 83, 2019.

[9] S. Twaha, J. Zhu, L. Maraaba, K. Huang, B. Li, and Y. Yan, "Maximum power point tracking control of a thermoelectric generation system using the extremum seeking control method," *Energy*, vol. 10, no. 12, 2017.

[10] J. M. Kwon, K. H. Nam, and K. H. Nam, "High-efficiency module-integrated photovoltaic power conditioning system," *IET Power Electronics*, vol. 2, no. 4, pp. 410–420, 2009.

[11] A. Ymeri and S. Mujović, "Optimal location and sizing of photovoltaic systems in order to reduce power losses and voltage drops in the distribution grid," *International Review of Electrical Engineering*, vol. 12, no. 6, pp. 498–504, 2017.

[12] M. Luthfansyah, S. Suyanto, and A. B. M. Bangura, "Evaluation and comparison of DC-DC power converter variations in solar panel systems using maximum power point tracking (MPPT) flower pollination algorithm (FPA) Control," in *Proceedings of the 1st International Conference on Renewable Energy Research and Challenge*, p. 606, Banyuwangi, Indonesia, 2020.

[13] M. Eltamaly Ali, "A novel musical chairs algorithm applied for MPPT of PV systems," *Renewable and Sustainable Energy Reviews*, vol. 3, p. 146, 2021.

[14] Q. Liu, S. Pang, and Y. Gao, "An intelligent dual-mode control strategy for extending MPPT operation interval of DFIG," *Electric Power Components and Systems*, vol. 48, no. 16-17, 2021.

[15] Ngo Van-Quang-Binh, L. Mohsen, A. Rabeh, J. Houssem, K. Ohshima, and K. Mehrdad, "Improved krill herd algorithm based sliding mode MPPT controller for variable step size P&O method in PV system under simultaneous change of irradiance and temperature," *Journal of the Franklin Institute*, vol. 358, no. 7, pp. 3491–3511, 2021.

[16] S. Kouro, J. I. Leon, D. Vinnikov et al., "Grid connected photovoltaic systems: an overview of recent research and emerging PV converter technology," *IEEE Industrial Electronics Magazine*, vol. 9, no. 1, pp. 47–61, 2015.

Hindawi
Complexity
Volume 2021, Article ID 5964034, 16 pages
https://doi.org/10.1155/2021/5964034

Research Article

Adaptive Finite-Time Fault-Tolerant Control for Half-Vehicle Active Suspension Systems with Output Constraints and Random Actuator Failures

Jie Lan[1,2] **and Tongyu Xu**[1]

[1]*College of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang 110000, Liaoning, China*
[2]*College of Science, Liaoning University of Technology, Jinzhou 121001, Liaoning, China*

Correspondence should be addressed to Tongyu Xu; xutongyu@syau.edu.cn

Received 29 April 2021; Revised 4 June 2021; Accepted 21 June 2021; Published 10 July 2021

Academic Editor: Rui Wang

The problem of adaptive finite-time fault-tolerant control (FTC) and output constraints for a class of uncertain nonlinear half-vehicle active suspension systems (ASSs) are investigated in this work. Markovian variables are used to denote in terms of different random actuators failures. In adaptive backstepping design procedure, barrier Lyapunov functions (BLFs) are adopted to constrain vertical motion and pitch motion to suppress the vibrations. Unknown functions and coefficients are approximated by the neural network (NN). Assisted by the stochastic practical finite-time theory and FTC theory, the proposed controller can ensure systems achieve stability in a finite time. Meanwhile, displacement and pitch angle in systems would not violate their maximum values, which imply both ride comfort and safety have been enhanced. In addition, all the signals in the closed-loop systems can be guaranteed to be semiglobal finite-time stable in probability (SGFSP). The simulation results illustrate the validity of the established scheme.

1. Introduction

With the development of modern industrial automation, vehicles play a more and more important role in people's production and life. Suspension as an important part, the damping effect to a great extent determines the comfort and safety of the automobile. Compared to the traditional passive and semiactive suspension systems, active suspension systems (ASSs) can provide better dynamic adjustment damping, potential road handling capacity, extreme ride comfort, and suspension deflection [1–3]. The design of complex mechanical engineering of ASSs had become a hot issue in the past two decades.

ASSs are often simplified into the full-vehicle model, half-vehicle model, and quarter-vehicle model. The actuators are parallel inserted to the components that provide external forces to increase or dissipate the energy of the ASSs and manage the tradeoffs between conflicting performance indicators. Many remarkable results were reported in [4–8] with the vertical motion of the quarter ASSs as the research topic. However, pitch motion was ignored that also directly affected the ride comfort and safety. Research on half-vehicle ASSs mainly focused on the discussions of pitch motion and vertical motion [9–17]. For ASSs, some inevitable uncertainties were in the design of controllers in [9–13], such as body mass, mass moment of inertia, and modeling uncertainties. But, these control methods [9–11] did not achieve good performance in estimating real valves. In [12, 13], they proposed adaptive control schemes by adding new leakage items to the update rules. In [14], the damping coefficient and spring stiffness of the tires were considered in the suspension as random uncertain parameters. The uncertain actuator was discussed and eliminated the influence by continuous-time homogeneous [15]. However, many progresses have been made for uncertain nonlinear ASSs, and a few studies were on the constraint of half-vehicle models.

The actual mechanical systems need to keep the output or states within certain ranges; otherwise, the system performance would degrade. The properties of prescribed performance control [17] and barrier Lyapunov functions (BLFs) provide effective methods to deal with constraints, and a large number of results have been obtained for output constraints of various nonlinear systems and practical systems [18–26]. The asymmetric BLFs were coped with the position constraint problem of the marine vessel [24]. Due to the limitation of physical factors in suspension fields, it is worth noting that [25, 26] reported the output and time-varying output constraint of vertical motion for the quarter ASSs. However, for the half-vehicle ASSs, considering above output constraints had been carried out.

Theoretically, the aforementioned works can only guarantee the desired system performance as time approaches infinity. However, the actual mechanical control should achieve the expected transient system performance. The design of finite-time control for nonlinear systems has attracted considerable attention. The finite-time Lyapunov stability theorem was first proposed in [27]. Based on this theory, the continuous finite-time control for nonlinear systems was proposed in [28–34], robotic manipulators in [28], switched systems in [32–34], and Markovian jump systems in [35, 36]. The concept of semiglobal practical finite-time stability (SGPFS) was proposed in different forms in [31–36]. The adaptive fuzzy finite-time control scheme of general uncertain nonlinear systems is discussed in [36]. Furthermore, Cai and Xiang expanded SGPFS to nonstrict nonlinear systems in [37]. Sui et al. expanded SGPFS for nontriangular stochastic nonlinear systems in [38]. The finite-time results of nonlinear quarter ASSs had been made in [39, 40]. The finite-time results of nonlinear half ASSs had been made in [41], but there was no constraint study on the individual states. For nonlinear strict feedback systems, both output constraint and finite-time control design had been completed in [42, 43]. However, for uncertain nonlinear half ASSs, there are few results on how to implement finite-time control associated with output constraints.

On the other hand, actuator failures are inevitable due to the influence of external environment, mechanical system failures, operation errors, and human factors in practical systems. These faults can seriously have an impact on system stability, degradation, and even catastrophic risks. Most of the above studies assumed that all actuators or sensors were in normal operation. Fault-tolerant control (FTC) strategies can compensate the faults and maintain acceptable system performance. Many methods to deal with actuator failures, such as the pseudoinverse method was in [44], model prediction method was in [45], and sliding control was in [46], by applying the adaptive backstepping techniques for linear systems in [47] and nonlinear strict feedback systems in [48–50]. Failures should be random. The actuator of states can switch between various modes in a random way. Given enough historical data, the states of the actuator can be modeled as Markov states in [51]. In the process, the failures of different actuators also meet the requirements of different Markov processes. Each actuator is independent and can fail at any sampling time. In [52], an adaptive fault compensation for a class of nonlinear uncertain systems with random actuator faults was studied and a random function to scale actuator faults by Markov correlation variables was proposed. In [53], random faults between different actuators in the half ASSs were considered for the first time. Motivated by the above observations and existing research results, this study proposes an adaptive NN finite-time FTC scheme for uncertain nonlinear half-vehicle ASSs with output constraints. The three main advantages of the proposed scheme can be listed as follows:

(1) Compared with existing adaptive FTC studies, the problem for half ASSs subject to infinite stochastic actuator failures and the states of multiactuators modeled by different Markovian processes has not received enough attention. Particularly, considering finite-time control, the additional correlation terms generated by the infinitesimal generator are handled by the stochastic finite-time control theorem.

(2) In comparison with existing constraints, it is asymmetry, which can restrain different outputs of displace and pitch angle more reasonable and reduce the vibration in uncertain nonlinear half ASSs. Moreover, the finite-time FTC control strategy can also enable the practical control systems to realize the transient system stability.

(3) In comparison with existing adaptive finite-time control, it considers a class of uncertain nonlinear half-ASSs subject to stochastic actuator failures. It should dispose random terms which makes the existing stability criteria in [53–55] are invalid. Concurrently, the asymmetric output constraints for different factors have been considered. The Lyapunov is proved SGPFS.

This work is organized as follows. In Section 2, the half active suspension systems and control objectives are shown. Section 3 presents the design procedures of the adaptive finite-time fault-tolerant controller designed based on stochastic Lyapunov function and zero dynamic. In Section 4, an example to show that the constructed method is effective. In Section 5, it demonstrates a conclusion about the results of this work and future work.

2. System Description and Preliminaries

2.1. Nonlinear Half-Vehicle Suspension Systems. Figure 1 shows a half-vehicle suspension model. M represents the mass of the vehicle body. I is the mass moment of inertia. m_f and m_r are the defined masses of front and rear wheels, respectively. D_c stands for the vertical displacement of the vehicle body. φ represents the pitch angle. D_f and D_r stand for the displacements of the front and rear vehicle body, respectively. D_1 and D_2 are the displacements of the front and rear wheels, respectively. D_{01} and D_{02} represent the road inputs of corresponding wheels. F_{sr}, F_{tr}, F_{sf}, and F_{tf} are defined as the forces produced by the related stiffness. F_{dr}, F_{br}, F_{df}, and F_{bf} are defined as the forces produced by the related dampers. u_f and u_r represent the control forces of the front and rear ASSs.

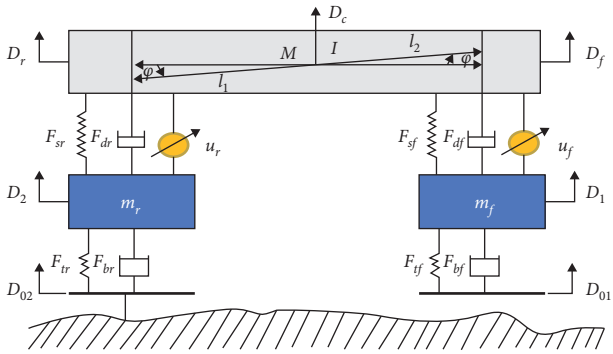


FIGURE 1: The nonlinear half-vehicle suspension model.

For half ASSs, the state space equations of vertical motion and pitch motion are shown in the following equation:

$$\begin{cases} M\ddot{D}_c + F_{\mathrm{sf}} + F_{\mathrm{df}} + F_{\mathrm{sr}} + F_{\mathrm{dr}} - u_z = 0, \\ I\ddot{\varphi} + l_2\left(F_{\mathrm{sf}} + F_{\mathrm{df}}\right) - l_1\left(F_{\mathrm{sr}} + F_{\mathrm{dr}}\right) - u_\varphi = 0, \\ m_f\ddot{D}_1 - F_{\mathrm{sf}} - F_{\mathrm{df}} + F_{\mathrm{tf}} + F_{\mathrm{bf}} + u_f = 0, \\ m_r\ddot{D}_2 - F_{\mathrm{sf}} - F_{\mathrm{dr}} + F_{\mathrm{tr}} + F_{\mathrm{br}} + u_r = 0, \end{cases} \tag{1}$$

where $u_z = u_r + u_f$ and $u_\varphi = l_2u_f - l_1u_r$.

The stochastic actuator failures are considered and described as follows:

$$\begin{aligned} u_f(t) &= g_1\left(r_1(t)\right)\Psi_f(t), \\ u_r(t) &= g_2\left(r_2(t)\right)\Psi_r(t), \end{aligned} \tag{2}$$

where $\Psi_f(t)$ and $\Psi_r(t)$ are the inputs of the front actuator and rear actuator, respectively. $r_1(t)$ and $r_2(t)$ are the independent irreducible right continuous homogeneous Markovian processes on the probability space (Ω, E, P), taking values in a finite set $S = \{1, 2, \ldots, N\}$ with generator matrix $\Pi = (\Upsilon_{pq})_{N\times N}$, where $\Upsilon_{pq} > 0$ is the transition rate from mode p to mode q if $p \neq q$, and $\Upsilon_{pp} = \sum_{q=1, p\neq q}^{N}\Upsilon_{pq}$.

In addition, $g_1(r_1(t))$ and $g_2(r_2(t))$ are the stochastic functions which represent the failure scaling factors for two different actuators and take values on the interval $[g_{1\min}, 1][g_{2\min}, 1]$, and $g_{1\min} > 0$, $g_{2\min} > 0$.

Remark 1. The values $g_1(r_1(t))$ and $g_2(r_2(t))$ meet as follows,

(1) When $g_1(r_1(t)) = 1$ or $g_2(r_2(t)) = 1$, the front actuator or the rear actuator is healthy

(2) When $g_1(r_1(t)) \in [g_{1\min}, 1)$ or $g_2(r_2(t)) \in [g_{2\min}, 1)$, there is partial failure of the corresponding actuator

Each actuator will switch randomly between above two states. The problem of FTC considered in the systems was assumed that once the actuator failed, it would keep the fault state for the rest operation. However, the failure may be intermittent, and the actuator may repetitiously fail with different failure modes. Meanwhile, the modes, times, and patterns of actuator failures are essentially random. Then, actuator failures in (2) are more complicated and practical.

In order to facilitate the design and analysis of the adaptive finite-time fault-tolerant control method, the state variables need to be defined as follows:

$$\begin{aligned} x_1 &= D_c, \\ x_2 &= \dot{D}_c, \\ x_3 &= \varphi, \\ x_4 &= \dot{\varphi}, \\ x_5 &= D_1, \\ x_6 &= \dot{D}_1, \\ x_7 &= D_2, \\ x_8 &= \dot{D}_2, \\ x &= \left[\, x_1 \;\; x_2 \;\; x_3 \;\; x_4 \;\; x_5 \;\; x_6 \;\; x_7 \;\; x_8 \,\right]^T. \end{aligned} \tag{3}$$

Then, it obtains

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = F_1 + \rho_1\left(g_1\left(r_1(t)\right)\Psi_f(t) + g_2\left(r_2(t)\right)\Psi_r(t)\right), \\ \dot{x}_3 = x_4, \\ \dot{x}_4 = F_2 + \rho_2\left(l_2g_1\left(r_1(t)\right)\Psi_f(t) - l_1g_2\left(r_2(t)\right)\Psi_r(t)\right), \end{cases} \tag{4}$$

$$\begin{cases} \dot{x}_5 = x_6, \\ \dot{x}_6 = \dfrac{1}{m_f}\left(F_{\mathrm{sf}} + F_{\mathrm{df}} - F_{\mathrm{tf}} - F_{\mathrm{bf}} - g_1\left(r_1(t)\right)\Psi_f(t)\right), \\ \dot{x}_7 = x_8, \\ \dot{x}_8 = \dfrac{1}{m_r}\left(F_{\mathrm{sr}} + F_{\mathrm{dr}} - F_{\mathrm{tr}} - F_{\mathrm{br}} - g_2\left(r_2(t)\right)\Psi_r(t)\right), \end{cases} \tag{5}$$

where $\rho_1 = 1/M$, $\rho_2 = 1/I$, $F_1 = \rho_1\left(-F_{\mathrm{sf}} - F_{\mathrm{df}} - F_{\mathrm{sr}} - F_{\mathrm{dr}}\right)$, and $F_2 = \rho_2\left(-l_2\left(F_{\mathrm{sf}} + F_{\mathrm{df}}\right) + l_1\left(F_{\mathrm{sr}} + F_{\mathrm{dr}}\right)\right)$.

The mass M and the mass moment of inertia I are uncertain due to the uncertainties of passengers and the load quality of cargoes. Therefore, F_1 and F_2 are the uncertainty and unknown functions.

2.2. Control Objectives. For different types of vehicles, improving ride comfort and safety is one of the most important requirements. Due to the hardware limitations, the following three requirements must be considered in the design and control process of ASSs.

First, the controllers u_f and u_r are subjected to random Markovian jumping failures in (2). The vertical displacement D_c and pitch angle motion φ are considered and limited in safety ranges. D_c and φ should be smaller which can largely enhance ride comfort.

Second, considering the driving safety, the wheels should make uninterrupted contact with the ground. It means that the dynamic tire load must be less than the static load, i.e.,

$$\begin{aligned} \left|F_{\mathrm{tf}} + F_{\mathrm{bf}}\right| &\le F_f, \\ \left|F_{\mathrm{tr}} + F_{\mathrm{br}}\right| &\le F_r, \end{aligned} \tag{6}$$

where $F_f + F_r = (M + m_f + m_r)g$, and $F_f(l_1 + l_2) = m_f g(l_1 + l_2) + Ml_1 g$.

Last, suspension constraints must be guaranteed because of the confined mechanical space, that is, suspension deflections should not exceed their maximum values.

$$\begin{aligned} \left|\Delta D_f\right| &\le \Delta\overline{D}_f, \\ \left|\Delta D_r\right| &\le \Delta\overline{D}_r, \end{aligned} \tag{7}$$

where $\Delta D_f = D_c + l_2 \sin\varphi - D_1$, and $\Delta D_r = D_c - l_1 \sin\varphi - D_2$.

There are some assumptions, definitions, and lemmas presented in order to facilitate the design and analysis of adaptive control scheme.

Assumption 1. For given constants $\underline{k}_{c1} > 0$, $\overline{k}_{c1} > 0$, $\underline{k}_{c2} > 0$, and $\overline{k}_{c2} > 0$, the following results on $y_{1\,d}$, $y_{3\,d}$, $\dot{y}_{1\,d}$, and $\dot{y}_{3\,d}$ hold, i.e.,

$$\begin{aligned} -\underline{B}_{11} \le y_{1\,d} &\le \overline{B}_{11}, \\ \left|\dot{y}_{1\,d}\right| &\le \overline{B}_{12}, \\ -\underline{B}_{31} \le y_{3\,d} &\le \overline{B}_{31}, \\ \left|\dot{y}_{3\,d}\right| &\le \overline{B}_{32}, \end{aligned} \tag{8}$$

where $\overline{B}_{11} < \overline{k}_{c1}$, $\underline{B}_{11} \le \underline{k}_{c1}$, $\overline{B}_{31} < \overline{k}_{c2}$, and $\underline{B}_{31} \le \underline{k}_{c2}$, for positive parameters $\underline{B}_{11}$, $\overline{B}_{11}$, $\underline{B}_{31}$, $\overline{B}_{31}$, $\overline{B}_{12}$, and $\overline{B}_{32}$ are selected and positive parameters.

Assumption 2. In general, the inputs of the road and their time derivative of D_{01} and D_{02} are limited. Therefore, existing positive parameters make sure the following inequality:

$$\begin{aligned} \left|D_{01}(t)\right| &\le \overline{D}_{11}, \\ \left|\dot{D}_{01}(t)\right| &\le \overline{D}_{12}, \\ \left|D_{02}(t)\right| &\le \overline{D}_{21}, \\ \left|\dot{D}_{02}(t)\right| &\le \overline{D}_{22}, \end{aligned} \tag{9}$$

where the parameters $\overline{D}_{11}$, $\overline{D}_{12}$, $\overline{D}_{21}$, and $\overline{D}_{22}$ are positive.

Definition 1 (see [35]). Consider the following nonlinear systems:

$$\begin{aligned} \dot{\zeta} &= g(\zeta, u), \\ g(0,0) &= 0, \quad \zeta \in \mathfrak{R}^n, u \in \mathfrak{R}^n, \end{aligned} \tag{10}$$

where ζ and u represent the state and input vectors, respectively. If any initial condition $\zeta(t_0)$ satisfies $\zeta(t_0) = \zeta_0$, for $t \ge t_0 + T$, the system (10) is semiglobal practical finite-time stable (SGPFS). In addition, the state vector $\zeta(t)$ satisfies $\|\zeta(t)\| \le \overline{\zeta}$, for $t \ge t_0 + T$, where $\overline{\zeta}$ is a positive parameter and a deposit time $T(\zeta_0, \overline{\zeta})$ satisfies $0 < T(\zeta_0, \overline{\zeta}) < \infty$.

Lemma 1 (see [36]). *For any $\gamma_i \in R$, $i = 1, 2, \ldots, n$, $0 < a < 1$, $0 < b < 2$, the following inequality can be constructed, namely,*

$$\begin{aligned} \left(\left|\gamma_1\right| + \left|\gamma_2\right| + \cdots + \left|\gamma_n\right|\right)^a &\le \left|\gamma_1\right|^a + \left|\gamma_2\right|^a + \cdots + \left|\gamma_n\right|^a, \\ \left(\left|\gamma_1\right|^2 + \left|\gamma_2\right|^2 + \cdots + \left|\gamma_n\right|^2\right)^b &\le \left(\left|\gamma_1\right|^a + \left|\gamma_2\right|^a + \cdots + \left|\gamma_n\right|^a\right)^2. \end{aligned} \tag{11}$$

Lemma 2 (see [37]). *Consider the Lyapunov function $V(\zeta)$ with stochastic terms, and the following inequality holds for a nonlinear system (10), namely,*

$$LV(\zeta) \le -\ell V^a(\zeta) + \hbar \tag{12}$$

where L is the differential operator, the constants ℓ and $\hbar$ are positive, and $0 < a < 1$. $\zeta \in R^n$ and $t \ge t_0$. Then, stochastic trajectory of (10) is SGPFS.

Lemma 3 (see [40]). *For any z and α are real variables, the following inequality can be constructed, namely,*

$$|z|^a|\alpha|^b \le \frac{a}{a+b}\mu|z|^{a+b} + \frac{b}{a+b}\mu^{-(a/b)}|\alpha|^{a+b}, \tag{13}$$

where the parameters a, b, and μ are positive.

Lemma 4 (see [23]). *For all $|\gamma_1| \le k_{b1}$, the following inequality can be constructed as*

$$\log\frac{1}{k_{b1}^2 - \gamma_1^2} \le \frac{\gamma_1^2}{k_{b1}^2 - \gamma_1^2}. \tag{14}$$

3. Adaptive Finite-Time Fault-Tolerant Controller Design Based on Stochastic Lyapunov Function and Zero Dynamic

For a class of uncertain nonlinear half ASSs, the universal approximation property of NN solves the uncertainty, and then, the outputs of vertical motion and pitch angle motion are constrained by using asymmetrical log-BLFs. The adaptive backstepping technique is used to address the functions with random generality variables generated by the infinitesimal generator due to random actuator failures. The Lyapunov stability is proved by the SGPFS and zero dynamic theorem.

In order to facilitate the design of controllers, there are some given coordinate transformations, i.e.,

$$\begin{aligned} \gamma_1 &= x_1 - y_{1\,d}, \\ \gamma_2 &= x_2 - \beta_1, \\ \gamma_3 &= x_3 - y_{3\,d}, \\ \gamma_4 &= x_4 - \beta_3, \end{aligned} \tag{15}$$

where β_1 and β_3 are the virtual control signals. γ_i for $1 \le i \le 4$ is the error variable. $y_{1\,d}$ and $y_{3\,d}$ are the desired vertical displacement and pitch angle, respectively.

3.1. Finite-Time Constraint Control Scheme Design for Vertical Motion and Pitch Angle Motion. Now, for the vertical motion of active suspension systems, more details on finite-time approach will be given in the next section.

Step 1. Select $V = V(x(t), r_1(t), r_2(t))$ as a Lyapunov function candidate in the following form:

$$V = \sum_{i=1,3}\left[\frac{1-q(\gamma_i)}{2}\log\frac{k_{ai}^2}{k_{ai}^2-\gamma_i^2} + \frac{q(\gamma_i)}{2}\log\frac{k_{bi}^2}{k_{bi}^2-\gamma_i^2}\right] + \sum_{i=1}^{2}\left(\frac{1}{2\rho_i}\gamma_{2i}^2 + \frac{1}{2\eta_i}\widetilde{\theta}_i^2 + \frac{1}{2\chi_i}\widetilde{g}_i^2(r_i(t))\right), \tag{16}$$

where $x(t) = (x_1(t), x_2(t), x_3(t), x_4(t))$, $\gamma_1 = (x_1, r_1(t), r_2(t))$, $\gamma_2 = (x_1, x_2, r_1(t), r_2(t))$, $\gamma_3 = (x_3, r_1(t), r_2(t))$, and $\gamma_4 = (x_3, x_4, r_1(t), r_2(t))$. The parameters η_i and χ_i are positive, and $\widetilde{\theta}_i$ and $\widetilde{g}_i(r_i(t))$ represent the errors. $\widehat{\theta}_i$ is the estimator of θ_i. $\widehat{g}_i(r_i(t))$ is the estimator of $g_i(r_i(t))$. Then, $\widetilde{\theta}_i = \theta_i - \widehat{\theta}_i$, $\widetilde{g}_i(r_i(t)) = g_i(r_i(t)) - \widehat{g}_i(r_i(t))$, $i = 1, 2$. It would make an assumption: $\forall t > 0$, $r_1(t) = p_1$, $r_2(t) = p_2$, and $p_1, p_2 \in S$.

Remark 2. For quarter ASSs, the study in [30] only showed the vehicle body's displacement constraint method by the symmetric BLFs. In contrast, we further study the pitch angle constraint of nonlinear uncertain half ASSs by asymmetric BLFs in this study.

In addition, the virtual controller signal is designed as

$$\beta_i = -\frac{c_i\gamma_i}{2}\left(\frac{1-q(\gamma_i)}{k_{ai}^2-\gamma_i^2} + \frac{q(\gamma_i)}{k_{bi}^2-\gamma_i^2}\right) + \dot{y}_{i\,d}, \tag{17}$$

where $i = 1, 3$. The parameter c_i is a design parameter, $c_i > 0$. k_{ai} and k_{bi} are the constraint bounded of corresponding variables, respectively. Moreover,

$$q(\cdot) = \begin{cases} 1, & \cdot \ge 0, \\ 0, & \cdot < 0. \end{cases} \tag{18}$$

Then, according to the infinitesimal generator of V, it gets

$$\mathrm{LV} = \sum_{i=1,3}\left[(1-q(\gamma_i))\frac{\gamma_i\dot{\gamma}_i}{k_{ai}^2-\gamma_i^2} + q(\gamma_i)\frac{\gamma_i\dot{\gamma}_i}{k_{bi}^2-\gamma_i^2}\right] + \sum_{i=1}^{2}\left(\frac{1}{\rho_i}\gamma_{2i}\dot{\gamma}_{2i} + \frac{1}{\eta_i}\widetilde{\theta}_i\dot{\widetilde{\theta}}_i + \frac{1}{\chi_i}\widetilde{g}_i(p_i)\dot{\widetilde{g}}_i(p_i)\right) + \sum_{q_1=1}^{N}\Upsilon_{p_1q_1}V_1 + \sum_{q_2=1}^{N}\Upsilon_{p_2q_2}V_2, \tag{19}$$

where $V_1 = (x(t), p_2, q_1)$, $V_2 = (x(t), p_1, q_2)$, and $\Upsilon_{p_iq_i} > 0$ is the transition rate, $i = 1, 2$.

Remark 3. $\sum_{q_1=1}^{N}\Upsilon_{p_1q_1}V_1$ and $\sum_{q_2=1}^{N}\Upsilon_{p_2q_2}V_2$ are the additional terms due to the involvement of Markovian variables $r_1(t)$ and $r_2(t)$, which do not exist in the determined situation. The extratransition rate-related terms appear in the infinitesimal generator of Lyapunov candidate function. These additions need further processing. It is a challenge that cannot be ignored in stability analysis.

From the previous definition, $i = 1, 2$, there are

$$\dot{\widetilde{g}}_i(r_i(t)) = -\dot{\widehat{g}}_i(r_i(t)), \qquad \dot{\widetilde{\theta}}_i = -\dot{\widehat{\theta}}_i. \tag{20}$$

Combining with (19) and (20) can obtain

$$\mathrm{LV} = \sum_{i=1,3}\gamma_i\left[\frac{1-q(\gamma_i)}{k_{ai}^2-\gamma_i^2} + \frac{q(\gamma_i)}{k_{bi}^2-\gamma_i^2}\right]\dot{\gamma}_i + \sum_{i=1}^{2}\frac{1}{\rho_i}\gamma_{2i}\dot{\gamma}_{2i} - \sum_{i=1}^{2}\left(\frac{1}{\chi_i}\widetilde{g}_i(p_i)\dot{\widehat{g}}_i(p_i) + \frac{1}{\eta_i}\widetilde{\theta}_i\dot{\widehat{\theta}}_i\right) + \sum_{q_1=1}^{N}\Upsilon_{p_1q_1}V_1 + \sum_{q_2=1}^{N}\Upsilon_{p_2q_2}V_2. \tag{21}$$

Furthermore, taking the derivative of γ_i, it yields

$$\dot{\gamma}_i = \gamma_{i+1} + \beta_i - \dot{y}_{i\,d}, \quad i = 1, 3. \tag{22}$$

Substituting (17) and (22) into (21), it can attain

$$\gamma_i\left[\frac{1-q(\gamma_i)}{k_{ai}^2-\gamma_i^2} + \frac{q(\gamma_i)}{k_{bi}^2-\gamma_i^2}\right]\dot{\gamma}_i = \lambda_i\left(\gamma_i\gamma_{i+1} - \frac{c_i\gamma_i^2}{2}\left(\frac{1-q(\gamma_i)}{k_{ai}^2-\gamma_i^2} + \frac{q(\gamma_i)}{k_{bi}^2-\gamma_i^2}\right)\right) = \lambda_i\gamma_i\gamma_{i+1} - c_i\frac{1-q(\gamma_i)}{2}\frac{\gamma_i^2}{k_{ai}^2-\gamma_i^2} - c_i\frac{q(\gamma_i)}{2}\frac{\gamma_i^2}{k_{bi}^2-\gamma_i^2}, \tag{23}$$

where $\lambda_i = ((1-q(\gamma_i))/k_{ai}^2-\gamma_i^2) + (q(\gamma_i)/(k_{bi}^2-\gamma_i^2))$, c_i is a positive design parameter, and $i = 1, 3$.

Based Lemma 3, let $z_i = (\gamma_i^2/2(k_{ai}^2-\gamma_i^2))$, $\alpha = 1$, $a = ((2n-1)/2n+1)$, $n \in \mathbb{N}$, $b = 1-a$, and $\mu = a^{(a/(1-a))}$ obtain

$$\left(\frac{\gamma_i^2}{2\left(k_{ai}^2-\gamma_i^2\right)}\right)^a \le (1-a)\mu+\frac{\gamma_i^2}{2\left(k_{ai}^2-\gamma_i^2\right)}. \tag{24}$$

Therefore, it gets

$$-\frac{\gamma_i^2}{2\left(k_{ai}^2-\gamma_i^2\right)} \le (1-a)\mu-\left(\frac{\gamma_i^2}{2\left(k_{ai}^2-\gamma_i^2\right)}\right)^a. \tag{25}$$

Constructed in the same way, it obtains

$$-\frac{(1-q(\gamma_i))\gamma_i^2}{2\left(k_{ai}^2-\gamma_i^2\right)} \le (1-a)\mu-(1-q(\gamma_i))^a\left(\frac{\gamma_i^2}{2\left(k_{ai}^2-\gamma_i^2\right)}\right)^a. \tag{26}$$

Similarly, it yields

$$\begin{aligned}-\frac{\gamma_i^2}{2\left(k_{bi}^2-\gamma_i^2\right)} &\le (1-a)\mu-\left(\frac{\gamma_i^2}{2\left(k_{bi}^2-\gamma_i^2\right)}\right)^a,\\ -\frac{q(\gamma_i)\gamma_i^2}{2\left(k_{bi}^2-\gamma_i^2\right)} &\le (1-a)-(q(\gamma_i))^a\left(\frac{\gamma_i^2}{2\left(k_{bi}^2-\gamma_i^2\right)}\right)^a.\end{aligned} \tag{27}$$

According to Lemma 4, for any constant a, $0<a<1$, the following inequality holds:

$$\begin{aligned}\left(\log\frac{1}{k_{ai}^2-\gamma_i^2}\right)^a &\le \left(\frac{\gamma_i^2}{k_{ai}^2-\gamma_i^2}\right)^a,\\ \left(\log\frac{1}{k_{bi}^2-\gamma_i^2}\right)^a &\le \left(\frac{\gamma_i^2}{k_{bi}^2-\gamma_i^2}\right)^a.\end{aligned} \tag{28}$$

Next, substituting (26)–(28) into (23), LV is expressed as

$$\begin{aligned}\mathrm{LV} = \sum_{i=1,3}\Bigg[&\lambda_i\gamma_i\gamma_{i+1}-c_i\left(\frac{1-q(\gamma_i)}{2}\log\frac{1}{k_{ai}^2-\gamma_i^2}\right)^a\\ &-c_i\left(\frac{q(\gamma_i)}{2}\log\frac{1}{k_{bi}^2-\gamma_i^2}\right)^a+\sum_{q_1=1}^{N}\Upsilon_{p_1q_1}V_1+\sum_{q_2=1}^{N}\Upsilon_{p_2q_2}V_2+\hbar_1\\ &+\sum_{i=1}^{2}\left(\frac{1}{\rho_i}\gamma_{2i}\dot\gamma_{2i}-\frac{1}{\chi_i}\tilde g_i(p_i)\dot{\hat g}_i(p_i)-\frac{1}{\eta_i}\tilde\theta_i\dot{\hat\theta}_i\right),\end{aligned} \tag{29}$$

where $\hbar_1=\sum_{i=1,3}2c_i(1-a)\mu$.

Taking the derivative of γ_2 and γ_4, they yield

$$\begin{aligned}\dot\gamma_2 &= F_1+\rho_1u_z-\dot\beta_1,\\ \dot\gamma_4 &= F_2+\rho_2u_\varphi-\dot\beta_3,\end{aligned} \tag{30}$$

where $\dot\beta_i=(\partial\beta_i/\partial x_i)\dot x_i+(\partial\beta_i/\partial y_{i\,d})\dot y_{i\,d}+(\partial\beta_i/\partial\dot y_{i\,d})\ddot y_{i\,d}$, $i=1,3$. β_i has been designed in (17). Then, $\sum_{i=1}^{2}(1/\rho_i)\gamma_{2i}\dot\gamma_{2i}$ gets

$$\begin{aligned}\frac{1}{\rho_1}\gamma_2\dot\gamma_2 &= \gamma_2(\overline{F}_1+u_z),\\ \frac{1}{\rho_2}\gamma_4\dot\gamma_4 &= \gamma_4\left(\overline{F}_2+u_\varphi\right),\end{aligned} \tag{31}$$

where the unknown functions $\overline{F}_1$ and $\overline{F}_2$ are denoted as

$$\begin{aligned}\overline{F}_1 &= \rho_1^{-1}F_1-\rho_1^{-1}\dot\beta_1,\\ \overline{F}_2 &= \rho_2^{-1}F_2-\rho_2^{-1}\dot\beta_3.\end{aligned} \tag{32}$$

Using the powerful approximating ability of NNs in [42, 43] for uncertain nonlinear systems, $\overline{F}_1$ and $\overline{F}_2$ can employ

$$\begin{aligned}\overline{F}_1 &= \Theta_1\Phi(Z_1)^T+\tau_1,\\ \overline{F}_2 &= \Theta_2\Phi(Z_2)^T+\tau_2,\end{aligned} \tag{33}$$

where $Z_1=[x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,\beta_1,\dot\beta_1,y_{1\,d},\dot y_{1\,d}]^T$, $Z_2=[x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,\beta_3,\dot\beta_3,y_{3\,d},\dot y_{3\,d}]^T$, and τ_i is the approximation error and the designed positive constant $\overline{\tau}_i$ that satisfies $|\tau_i|\le\overline{\tau}_i$. More remarkable, $0<\Theta_{1i}\Phi(Z_1)<1$ and $0<\Theta_{2i}\Phi(Z_2)<1$. Furthermore, we can draw a conclusion $0<\Phi_1^T(Z_1)\Phi_1(Z_1)\le N_1$ and $0<\Phi_2^T(Z_2)\Phi_2(Z_2)\le N_2$, where N_1 and N_2 are the corresponding numbers of NNs nodes.

Then, applying Young's inequality, it can get the following:

$$\begin{aligned}\gamma_{2i}\Theta_i\Phi(Z_i)^T &\le \frac{\gamma_{2i}^2\|\Theta_i\|^2\Phi(Z_i)^T\Phi(Z_i)}{2\delta_i^2}+\frac{\delta_i^2}{2}\\ &= \frac{\gamma_{2i}^2\theta_i}{2\delta_i^2}+\frac{\delta_i^2}{2},\end{aligned} \tag{34}$$

$$\gamma_{2i}\tau_i \le \frac{1}{2}\gamma_{2i}^2+\frac{1}{2}\tau_i^2 \le \frac{1}{2}\gamma_{2i}^2+\frac{1}{2}\overline{\tau}_i^2, \tag{35}$$

where $\theta_i=N_i\|\Theta_i\|^2$ and the design parameter δ_i is positive.

Substituting (33)–(35) into (31), it obtains

$$\begin{aligned}\sum_{i=1}^{2}\frac{1}{\rho_i}\gamma_{2i}\dot\gamma_{2i} \le{}& \gamma_2\left(\frac{\gamma_2\theta_1}{2\delta_1^2}+\frac{\gamma_2}{2}+u_z\right)+\gamma_4\left(\frac{\gamma_4\theta_2}{2\delta_2^2}+\frac{\gamma_4}{2}+u_\varphi\right)\\ &+\sum_{i=1}^{2}\left(\frac{\delta_i^2}{2}+\frac{\overline{\tau}_i^2}{2}\right).\end{aligned} \tag{36}$$

Substituting (36) into (29), it can obtain

$$\begin{aligned}\mathrm{LV} = \sum_{i=1,3}\Bigg[&\lambda_i\gamma_i\gamma_{i+1}-c_i\left(\frac{1-q(\gamma_i)}{2}\log\frac{1}{k_{ai}^2-\gamma_i^2}\right)^a\\ &-c_i\left(\frac{q(\gamma_i)}{2}\log\frac{1}{k_{bi}^2-\gamma_i^2}\right)^a+\gamma_2\left(\frac{\gamma_2\theta_1}{2\delta_1^2}+\frac{\gamma_2}{2}+u_z\right)\\ &+\gamma_4\left(\frac{\gamma_4\theta_2}{2\delta_2^2}+\frac{\gamma_4}{2}+u_\varphi\right)\\ &-\sum_{i=1}^{2}\left(\frac{\tilde g_i(p_i)\dot{\hat g}_i(p_i)}{\chi_i}+\frac{\tilde\theta_i\dot{\hat\theta}_i}{\eta_i}\right)+\sum_{q_1=1}^{N}\Upsilon_{p_1q_1}V_1+\sum_{q_2=1}^{N}\Upsilon_{p_2q_2}V_2+\hbar_2,\end{aligned} \tag{37}$$

where $\hbar_2=\hbar_1+\sum_{i=1}^{2}((\delta_i^2/2)+(\overline{\tau}_i^2/2))$.

Design the control input u_z and u_φ that are subjects to random actuator faults as

$$u_z = -\lambda_1\gamma_1 - \frac{\gamma_2}{2} - \frac{\gamma_2\widehat{\theta}_1}{2\delta_1^2} - c_2\gamma_2^{2a-1},$$
$$u_\varphi = -\lambda_3\gamma_3 - \frac{\gamma_4}{2} - \frac{\gamma_3\widehat{\theta}_2}{2\delta_3^2} - c_4\gamma_4^{2a-1}, \tag{38}$$

where the parameters c_2 and c_4 are positive.

The adaptive law is established as

$$\dot{\widehat{\theta}}_i = \frac{\eta_i}{2\delta_i^2}\gamma_{2i}^2 - \kappa_i\widehat{\theta}_i, \quad i = 1, 2, \tag{39}$$

where the designed parameter κ_i is positive.

Based on (1), (2), and (39), u_f and u_r are shown as

$$u_f = \frac{l_1u_z + u_\varphi}{l_1 + l_2},$$
$$u_r = \frac{l_2u_z - u_\varphi}{l_1 + l_2},$$
$$\Psi_f(t) = \frac{1}{\widehat{g}_1(l_1 + l_2)}\left[l_1\left(-\lambda_1\gamma_1 - \frac{\gamma_2}{2} - \frac{\gamma_2\widehat{\theta}_1}{2\delta_1^2} - c_2\gamma_2^{2a-1}\right) - \lambda_3\gamma_3 - \frac{\gamma_4}{2} - \frac{\gamma_3\widehat{\theta}_2}{2\delta_3^2} - c_4\gamma_4^{2a-1}\right], \tag{40}$$
$$\Psi_r(t) = \frac{1}{\widehat{g}_2(l_1 + l_2)}\left[l_2\left(-\lambda_1\gamma_1 - \frac{\gamma_2}{2} - \frac{\gamma_2\widehat{\theta}_1}{2\delta_1^2} - c_2\gamma_2^{2a-1}\right) + \lambda_3\gamma_3 + \frac{\gamma_4}{2} + \frac{\gamma_3\widehat{\theta}_2}{2\delta_3^2} + c_4\gamma_4^{2a-1}\right].$$

The updating laws of estimated parameters are established:

$$\dot{\widehat{g}}_1(p_1) = \mathrm{proj}^{(\gamma_2 + l_1\gamma_4)\chi_1\psi_f(t) - 2c_{g_1}\widehat{g}_1(p_1)}_{\widehat{g}_1(p_1)} \begin{cases} 0, & \widehat{g}_1(p_1) = g_{1\min}, \\ (\gamma_2 + l_1\gamma_4)\chi_1\psi_f(t) - 2c_{g_1}\widehat{g}_1(p_1), \widehat{g}_1(p_1), & \text{others}, \end{cases}$$
$$\dot{\widehat{g}}_2(p_2) = \mathrm{proj}^{(\gamma_2 - l_2\gamma_4)\chi_2\psi_r(t) - 2c_{g_2}\widehat{g}_2(p_2)}_{\widehat{g}_2(p_2)} \begin{cases} 0, & \widehat{g}_2(p_2) = g_{2\min}, \\ (\gamma_2 - l_1\gamma_4)\chi_2\psi_r(t) - 2c_{g_2}\widehat{g}_2(p_2), \widehat{g}_2(p_2), & \text{others}, \end{cases} \tag{41}$$

where proj represents the projection operator in order to avoid singular values in the denominator.

Remark 4. The adaptive estimated parameters $\dot{\widehat{g}}_1(p_1)$ and $\dot{\widehat{g}}_2(p_2)$ appear in the denominators in (41), which may cause the controllers to fail. Construct the projection form to avoid this situation.

Substituting (39)–(41) into (37), LV can be expressed as

$$\begin{aligned} \mathrm{LV} \le & -\sum_{i=1,3} c_i\left[\left(\frac{1 - q(\gamma_i)}{2}\log\frac{1}{k_{ai}^2 - \gamma_i^2}\right)^a + \left(\frac{q(\gamma_i)}{2}\log\frac{1}{k_{bi}^2 - \gamma_i^2}\right)^a\right] \\ & - \sum_{i=1}^{2} c_{2i}\gamma_{2i}^{2a} + \sum_{i=1}^{2}\frac{c_{g_i}\widetilde{g}_i(p_i)\widehat{g}_i(p_i)}{\chi_i} + \frac{\kappa_i\widetilde{\theta}_i\widehat{\theta}_i}{\eta_i} \\ & + \sum_{q_1=1}^{N}\Upsilon_{p_1q_1}V_1 + \sum_{q_2=1}^{N}\Upsilon_{p_2q_2}V_2 + \hbar_2. \end{aligned} \tag{42}$$

By Young's inequality $(i = 1, 2)$, they can obtain

$$\widetilde{\theta}_i\widehat{\theta}_i \le \frac{1}{2}\widetilde{\theta}_i^2 + \frac{1}{2}\theta_i^2,$$
$$\widetilde{g}_i(p_i)\widehat{g}_i(p_i) \le -\frac{1}{2}\widetilde{g}_i^2(p_i) + \frac{1}{2}\widehat{g}_i^2(p_i). \tag{43}$$

LV can be rewritten as

$$\begin{aligned} \mathrm{LV} \le & -\sum_{i=1,3} c_i\left[\left(\frac{1 - q(\gamma_i)}{2}\log\frac{1}{k_{ai}^2 - \gamma_i^2}\right)^a + \left(\frac{q(\gamma_i)}{2}\log\frac{1}{k_{bi}^2 - \gamma_i^2}\right)^a\right] \\ & - \sum_{i=1}^{2}\left[c_{2i}\gamma_{2i}^{2a} + \frac{c_{g_i}\widetilde{g}_i^2(p_i)}{2\chi_i} + \frac{\kappa_i\widetilde{\theta}_i^2}{2\eta_i}\right] \\ & + \sum_{q_1=1}^{N}\Upsilon_{p_1q_1}V_1 + \sum_{q_2=1}^{N}\Upsilon_{p_2q_2}V_2 + \hbar_3, \end{aligned} \tag{44}$$

where $\hbar_3 = \hbar_2 + \sum_{i=1}^{2}((\kappa_i\theta_i^2/2\eta_i) + (c_{g_i}g_i^2(p_i)/2\chi_i))$.

Based on Lemma 3, let $z = (\widetilde{\theta}_i^2/2\eta_i)a = (2n-1)/(2n+1)$, $n \in \mathbb{N}$, $b = 1 - a$, $\alpha = 1$, and $\mu = a^{(a/(1-a))}$, one has

$$\left(\frac{\widetilde{\theta}_i^2}{2\eta_i}\right)^a \le (1-a)\mu + \frac{\widetilde{\theta}_i^2}{2\eta_i}. \tag{45}$$

Furthermore, we can obtain

$$-\frac{\widetilde{\theta}_i^2}{2\eta_i} \le (1-a)\mu - \left(\frac{\widetilde{\theta}_i^2}{2\eta_i}\right)^a. \tag{46}$$

Applying the inequality (44) and (46) can be expressed as

$$\begin{aligned}\mathrm{LV} \le & -\sum_{i=1,3} c_i\left[\left(\frac{1-q(\gamma_i)}{2}\log\frac{1}{k_{ai}^2-\gamma_i^2}\right)^a + \left(\frac{q(\gamma_i)}{2}\log\frac{1}{k_{bi}^2-\gamma_i^2}\right)^a\right] - \sum_{i=1}^{2}\left[c_{2i}\gamma_{2i}^{2a} + \frac{c_{g_i}\widetilde{g}_i^2(p_i)}{2\chi_i} + \kappa_i\left(\frac{\widetilde{\theta}_i^2}{2\eta_i}\right)^a\right] \\ & + \sum_{q_1=1}^{N}\Upsilon_{p_1q_1}V_1 + \sum_{q_2=1}^{N}\Upsilon_{p_2q_2}V_2 + \hbar,\end{aligned} \tag{47}$$

where $\hbar = \hbar_3 + (1-a)\mu$.

In addition, the number of passengers and the load conditions of the vehicle are limited. The mass M is bounded which means ρ_i is bounded, namely,

$$\underline{\rho}_i \le \rho_i \le \overline{\rho}_i, \quad i = 1, 2. \tag{48}$$

Based on (48),

$$\begin{aligned}\mathrm{LV} \le & -\sum_{i=1,3} c_i\left[\left(\frac{1-q(\gamma_i)}{2}\log\frac{1}{k_{ai}^2-\gamma_i^2}\right)^a + \left(\frac{q(\gamma_i)}{2}\log\frac{1}{k_{bi}^2-\gamma_i^2}\right)^a\right] \\ & -\sum_{i=1}^{2}\left[c_{2i}\left(2\underline{\rho}_i\right)^a\left(\frac{\gamma_{2i}^{2a}}{2\underline{\rho}_i}\right)^a + \frac{c_{g_i}\widetilde{g}_i^2(p_i)}{2\chi_i} + \kappa_i\left(\frac{\widetilde{\theta}_i^2}{2\eta_i}\right)^a\right] \\ & + \sum_{q_1=1}^{N}\Upsilon_{p_1q_1}V_1 + \sum_{q_2=1}^{N}\Upsilon_{p_2q_2}V_2 + \hbar.\end{aligned} \tag{49}$$

From (16), (49) can be rewritten as

$$\mathrm{LV} \le -\ell V^a + \hbar + \sum_{q_1=1}^{N}\Upsilon_{p_1q_1}V_1 + \sum_{q_2=1}^{N}\Upsilon_{p_2q_2}V_2, \tag{50}$$

where $\ell = \min\left\{-c_2(2\underline{\rho}_1)^a, c_4(2\underline{\rho}_2)^a, c_1, c_3, \kappa_1, \kappa_2, c_{g_1}, c_{g_2}\right\}$.

The following will apply the discrete expectation to address the random terms.

$$\lim_{\|x\|\longrightarrow\infty}\inf_{t\ge t_0} V = \infty. \tag{51}$$

There would be a B, no matter how big it is; once more than $t > t_l$, it would exceed B. From (51) in $[t_0, t_l]$, $\|x(t)\| \le B$. (50) is bounded in $[t_0, t_l]$.

Then, according to property of discrete expectations, it yields

$$\begin{aligned}E(\mathrm{LV}) &\le \sum_{p_1=1}^{N}\sum_{p_2=1}^{N} E(\mathrm{LV}_l)\pi_{p_1}\pi_{p_2} \\ &\le -\ell\sum_{p_1=1}^{N}\sum_{p_2=1}^{N} E(\mathrm{LV}_l^a)\pi_{p_1}\pi_{p_2} + \hbar \\ &\quad + \sum_{p_1=1}^{N}\sum_{p_2=1}^{N} E\left(\sum_{q_1=1}^{N}\Upsilon_{p_1q_1}V_{1l} + \sum_{q_2=1}^{N}\Upsilon_{p_2q_2}V_{2l}\right)\pi_{p_1}\pi_{p_2} \\ &\le -\ell\sum_{p_1=1}^{N}\sum_{p_2=1}^{N} E(\mathrm{LV}_l^a) + \hbar + N\left(\sum_{k=1}^{2}\max_{p_k,q_k=1}^{N}\left\{\gamma_{p_kq_k}\frac{\pi_{p_k}}{\pi_{q_k}}\right\}\right)EV_l^a \le -(\ell-\ell_l)EV_l^a + \hbar,\end{aligned} \tag{52}$$

where $V_l = V(x(t_l), r_1(t_l), r_2(t_l))$, $V_{1l} = (x(t_l), p_2, q_1)$, $V_{2l} = (x(t_l), p_1, q_2)$, and $V_l^a = V^a(x(t_l), r_1(t_l), r_2(t_l))$. The stability distribution law of Markovian variable $r_i(t_l)$ is $\pi_{p_i} = (\pi_1, \pi_2, \ldots, \pi_N)$. $\sum_{j=1}^{N} \pi_j = 1$, and $\pi_j > 0$. $\ell_l = N(\sum_{k=1}^{2} \max_{p_k,q_k=1}^{N} \{\gamma_{p_k q_k}(\pi_{p_k}/\pi_{q_k})\})$. Various parameters associated with the selection need to be satisfied, $(\ell - \ell_l) > 0$.

$$\begin{aligned} EV_l^a &= V_{t_0}^a + E\int_{t_0}^{t_l} e^{(\ell-\ell_l)s} \mathrm{L}V_s^a \mathrm{d}s \\ &\le V_{t_0}^a + \hbar(t_l - t_0) - (\ell - \ell_l)\int_{t_0}^{t_l} EV_s^a \mathrm{d}s \\ &\le V_{t_0}^a + \hbar(t_l - t_0), \end{aligned} \tag{53}$$

where $V_{t_0}^a = V^a(x(t_0), r_1(t_0), r_2(t_0))$.

For $\forall x(t_0)$ and $r_1(t_0), r_2(t_0) \in S$, there is a unique solution $x(t)$ in the closed-loop systems.

$$\begin{aligned} &E\left(e^{(\ell-\ell_l)t_l} V_l^a\right) - E\left(e^{(\ell-\ell_l)t_0} V_{t_0}^a\right) \\ &= E\left(\int_{t_0}^{t_l} e^{(\ell-\ell_l)s} \mathrm{L}V_s^a \mathrm{d}s\right) + (\ell - \ell_l)E\left(\int_{t_0}^{t_l} e^{(\ell-\ell_l)s} \mathrm{L}V_s^a \mathrm{d}s\right) \\ &\le -(\ell - \ell_l)E\left(\int_{t_0}^{t_l} e^{(\ell-\ell_l)s} \mathrm{L}V_s^a \mathrm{d}s\right) + \int_{t_0}^{t_l} \hbar e^{(\ell-\ell_l)s} \mathrm{d}s \\ &\quad + (\ell - \ell_l)E\left(\int_{t_0}^{t_l} e^{(\ell-\ell_l)s} V_s^a \mathrm{d}s\right) \\ &E\left(e^{(\ell-\ell_l)\ell_l} V_l^a\right) \le e^{(\ell-\ell_l)t_0} V_{t_0}^a + \frac{\hbar}{\ell - \ell_l}\left(e^{(\ell-\ell_l)t} - e^{(\ell-\ell_l)t_0}\right) \\ &E(V_l^a) \le e^{-(\ell-\ell_l)(t-t_0)} V_{t_0}^a + \frac{\hbar}{\ell - \ell_l}\left(1 - e^{-(\ell-\ell_l)(t-t_0)}\right) \\ &\le V_{t_0}^a + \frac{\hbar}{\ell - \ell_l}. \end{aligned} \tag{54}$$

From the Lemma 2, it can be concluded that all the signals can be guaranteed SGPFS. Meanwhile, the constraints of two different output factors are bounded.

Theorem 1. *The finite-time FTC control can be implemented by designing the virtual signal (17), the controller in (40), and the updating law (39) and (41). Meantime, by selecting appropriate design parameters, the tracking error of the systems can be arbitrarily small in a finite-time.*

Proof. Through Lemma 2 in [37] and inequality (54), we can acquire that the trajectories of the all signals that satisfy $V^a(\gamma(t), \widetilde{\theta}(t)) \le (\hbar/(1-\theta_0)\ell_{12})$, for any $t \ge T_{\text{reach}}$, in finite time, where T_{reach} is defined as $T_{\text{reach}} = (1/(1-\theta_0)\ell_{12})(V^{1-a}(\gamma(0), \theta_1(0))) - (\hbar/(1-\theta_0)\ell_{12})^{(1-a)/a}$ with $\gamma(0) = [\gamma_1(0), \gamma_2(0)]^T$. In other words, all signals of the studied closed system are SGPFS.

Furthermore, the following result can be obtained, $i = 1, 3$. $|x_i - y_{i\,d}| \le k_i[1 - e^{-2(\hbar/(1-\theta_0)\ell_{12})^{1/a}}]^{(1/2)}$, where $k_i = \min\{k_{ai}, k_{bi}\}$, In addition, the tracking error γ_i remains in a small neighborhood of the origin by finite-time control.

When $q(\gamma_i) = 1$ and $\gamma_i \ge 0$, we can get $\gamma_i \le k_{bi}[1 - e^{-2(\hbar/(1-\theta_0)\ell_{12})^{1/a}}]^{(1/2)} \le k_{bi}$. Thus, when $q(\gamma_i) = 0$ and $\gamma_i < 0$, it can obtain $\gamma_i \ge -k_{ai}[1 - e^{-2(\hbar/(1-\theta_0)\ell_{12})^{1/a}}]^{(1/2)} \ge -k_{ai}$. By using $-\underline{B}_{i1} \le y_{i\,d} \le \overline{B}_{i1}$, it yields $-\underline{k}_{ci} \le -k_{ai} + y_{i\,d} \le x_i \le k_{bi} + y_{i\,d} \le \overline{k}_{ci}$. In short, x_i is constrained in its limit. □

3.2. Zero Dynamics and Performance Analysis. Through the analysis in Section 3.1, the signals x_i and $\widehat{\theta}_j$ $(i = 1, 2, 3, 4, j = 1, 2)$ are bounded. Then, we are going to prove the zero dynamics that consist of the other four states x_5, x_6, x_7, and x_8.

Therefore, substituting (38)–(41) into (5), let $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$. It yields the following form:

$$\dot{Z} = AZ + B_1 D_1 + B_0 D_0 + W, \tag{55}$$

where

$$Z = [x_5, x_6, x_7, x_8]^T,$$

$$D_0 = [D_{01}, \dot{D}_{01}, D_{02}, \dot{D}_{02}]^T,$$

$$D_1 = [D_f - D_1, \dot{D}_f - \dot{D}_1, D_r - D_2, \dot{D}_r - \dot{D}_2]^T,$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{-k_{f2}}{m_f} & \frac{-b_{f2}}{m_f} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{-k_{r2}}{m_r} & \frac{-b_{r2}}{m_r} \end{bmatrix},$$

$$B_0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \frac{k_{f2}}{m_f} & \frac{b_{f2}}{m_f} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{k_{r2}}{m_r} & \frac{k_{r2}}{m_r} \end{bmatrix},$$

$$B_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \dfrac{\widetilde{g}_1 k_{f2}}{m_f \widehat{g}_1} & \dfrac{\widetilde{g}_1 b_{f2}}{m_f \widehat{g}_1} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \dfrac{\widetilde{g}_2 k_{r2}}{m_r \widehat{g}_2} & \dfrac{\widetilde{g}_2 b_{r2}}{m_r \widehat{g}_2} \end{bmatrix},$$

$$w = \begin{bmatrix} 0 \\ \dfrac{-l_2 w_1 - w_2}{m_f (l_1 + l_2)} \\ 0 \\ \dfrac{-l_2 w_1 + w_2}{m_r (l_1 + l_2)} \end{bmatrix}, \tag{56}$$

$$w_1 = -\lambda_1 \gamma_1 - \frac{1}{2}\gamma_2 - \frac{\gamma_2}{2\delta_1^2}\widehat{\theta}_1 - c_2 \gamma_2^{2l-1},$$

$$w_2 = -\lambda_3 \gamma_3 - \frac{1}{2}\gamma_4 - \frac{\gamma_4}{2\delta_3^2}\widehat{\theta}_2 - c_4 \gamma_4^{2l-1}.$$

Notable is, due to the boundedness of x_i and $\widehat{\theta}_j (i = 1, 2, 3, 4, j = 1, 2)$, w_1 and w_2 are bounded, namely, $\|w_i\| \le \varpi_i$ for the positive parameter $\varpi_i (i = 1, 2)$. Furthermore, it means that W satisfies $\|W\| \le \overline{W}$, for the parameter $\overline{W} > 0$. Meanwhile, according to Assumptions 1 and 2, D satisfies $\|D\| \le \overline{D}$, for the parameter $\overline{D} > 0$. Choosing the positive matrixes P and Q can ensure the following equality holds, namely,

$$A^T P + P^T A = -Q. \tag{57}$$

Define

$$V_Z = Z^T P Z, \tag{58}$$

where $V_Z = V_Z(Z, p_1, p_2)$, and P is a positive matrix.

According to the infinitesimal generator in [50], (58) can be calculated at the mode of $(r_1(t), r_2(t)) = (p_1, p_2)$ as

$$\mathrm{LV}_Z = \dot{Z}^T P Z + Z^T P \dot{Z} + \sum_{q_1=1}^{N} \Upsilon_{p_1 q_1} Z_1 + \sum_{q_2=1}^{N} \Upsilon_{p_2 q_2} Z_2, \tag{59}$$

where $Z_1 = Z_1(Z, p_2, q_1)$, and $Z_2 = Z_2(Z, p_1, q_2)$.

$$\begin{aligned} \mathrm{LV}_Z &= (AZ + B_1 D_1 + B_0 D_0 + W)^T P Z, \\ &Z^T P (AZ + B_1 D_1 + B_0 D_0 + W) + \sum_{q_1=1}^{N} \Upsilon_{p_1 q_1} Z_1 + \sum_{q_2=1}^{N} \Upsilon_{p_2 q_2} Z_2 \\ &= Z^T \left(A^T P + P^T A\right) Z + 2 Z^T P B_1 D_1 + 2 Z^T P B_0 D_0 + 2 Z^T P W, \\ &\quad + \sum_{q_1=1}^{N} \Upsilon_{p_1 q_1} Z_1 + \sum_{q_2=1}^{N} \Upsilon_{p_2 q_2} Z_2. \end{aligned} \tag{60}$$

Based on Young's inequality, it obtains

$$\begin{aligned} 2 Z^T P B_0 D_0 &\le \frac{\lambda_{\max}\left(P B_0^T B_0 P\right)}{\vartheta_0} \|Z\|^2 + \vartheta_0 \overline{D}_0^2, \\ 2 Z^T P B_1 D_1 &\le \frac{\lambda_{\max}\left(P B_1^T B_1 P\right)}{\vartheta_1} \|Z\|^2 + \vartheta_1 \overline{D}_1^2, \\ 2 Z^T P W &\le \frac{\lambda_{\max}(PP)}{\vartheta_2} \|Z\|^2 + \vartheta_2 \overline{W}^2, \end{aligned} \tag{61}$$

where the parameters ϑ_0, ϑ_1, and ϑ_2 are positive.

Combining (57) and (61), (60) can be rewritten as

$$\begin{aligned} \mathrm{LV}_Z &= -Z^T Q Z + \frac{\lambda_{\max}\left(P B_0^T B_0 P\right)}{\vartheta_0} \|Z\|^2 + \frac{\lambda_{\max}\left(P B_1^T B_1 P\right)}{\vartheta_1} \|Z\|^2 \\ &\quad + \frac{\lambda_{\max}(PP)}{\vartheta_2} \|Z\|^2 + \hbar_Z + \sum_{q_1=1}^{N} \Upsilon_{p_1 q_1} Z_1 + \sum_{q_2=1}^{N} \Upsilon_{p_2 q_2} Z_2 \\ &\le -Z^T \left(Q - \frac{\lambda_{\max}\left(P B_0^T B_0 P\right)}{\vartheta_0} - \frac{\lambda_{\max}\left(P B_1^T B_1 P\right)}{\vartheta_1} - \frac{\lambda_{\max}(PP)}{\vartheta_2} \right) Z + \hbar_Z + \sum_{q_1=1}^{N} \Upsilon_{p_1 q_1} Z_1 + \sum_{q_2=1}^{N} \Upsilon_{p_2 q_2} Z_2 \\ LV_Z &\le -\ell_Z V_Z + \hbar_Z + \sum_{q_1=1}^{N} \Upsilon_{p_1 q_1} Z_1 + \sum_{q_2=1}^{N} \Upsilon_{p_2 q_2} Z_2, \end{aligned} \tag{62}$$

where

$$\hbar_Z = \vartheta_0 \overline{D}_0^2 + \vartheta_1 \overline{D}_1^2 + \vartheta_2 \overline{W}^2,$$
$$\ell_Z = \lambda_{\min} P^{-(1/2)} Q P^{-(1/2)} - \frac{\lambda_{\max}(P)}{\vartheta_2} - \frac{\lambda_{\max}\left(P^{(1/2)} B_0^T B_0 P^{(1/2)}\right)}{\vartheta_0} - \frac{\lambda_{\max}\left(P^{(1/2)} B_1^T B_1 P^{(1/2)}\right)}{\vartheta_1}. \tag{63}$$

Then, same as the above discrete expectations, (62) yields

$$\mathrm{EV}_Z \le e^{-\ell_0 (t-t_0)} V_{Z_0} + \frac{\hbar_Z}{\ell_0}\left(1 - e^{-\ell_0 (t-t_0)}\right), \tag{64}$$

where $\ell_0 = \ell_Z - \ell_l$, and $V_{Z_0}(Z(t_0), r_1(t_0), r_2(t_0))$.

According to Lemma 2, all the signals of the zero dynamics are bounded.

Now, the control objectives will be proved. The dynamic tire loads are given as

$$\left|F_{\mathrm{tf}} + F_{\mathrm{bf}}\right| \le \left(k_{f_2} + b_{f_2}\right)\left(\frac{\left(V_{Z_0} + \hbar_3/\ell_Z\right)}{\lambda_{\min}(P)}\right)^{(1/2)} + k_{f_2}\overline{D}_{11} + b_{f_2}\overline{D}_{12},$$
$$\left|F_{\mathrm{tr}} + F_{\mathrm{br}}\right| \le \left(k_{r_2} + b_{r_2}\right)\left(\frac{\left(V_{Z_0} + \hbar_3/\ell_Z\right)}{\lambda_{\min}(P)}\right)^{(1/2)} + k_{r_2}\overline{D}_{21} + b_{r_2}\overline{D}_{22}. \tag{65}$$

Thus, selecting appropriate parameters ϑ_1 and ϑ_2 can ensure $(k_{f_2} + b_{f_2})\vartheta_3 + k_{f_2}\overline{D}_{11} + b_{f_2}\overline{D}_{12} \le F_f$, $(k_{r_2} + b_{r_2})\vartheta_3 + k_{r_2}\overline{D}_{21} + b_{r_2}\overline{D}_{22} \le F_r$ hold, that is, (6) holds.

Finally, we are going to learn about the suspension deflection performances. According to the analysis in Section 3.1 and 3.2, one has

$$\left|\Delta\overline{D}_f\right| \le x_1 + l_2 \sin x_3 + x_5 \le x_1 + l_2 x_3 + x_5 \le k_{c_1} + l_2 k_{c_2} + \left(\frac{\left(V_{Z_0} + \hbar_3/\ell_Z\right)}{\lambda_{\min}(P)}\right)^{(1/2)},$$
$$\left|\Delta\overline{D}_r\right| \le x_1 + l_2 \sin x_3 + x_7 \le x_1 + l_2 x_3 + x_7 \le k_{c_1} + l_2 k_{c_2} + \left(\frac{\left(V_{Z_0} + \hbar_3/\ell_Z\right)}{\lambda_{\min}(P)}\right)^{(1/2)}. \tag{66}$$

Thus, let $k_{c_1} + l_2 k_{c_2} + ((V_{Z_0} + \hbar_3/\ell_Z)/\lambda_{\min}(P))^{(1/2)} \le \Delta\overline{D}_f$ and $k_{c_1} + l_2 k_{c_2} + ((V_{Z_0} + \hbar_3/\ell_Z)/\lambda_{\min}(P))^{(1/2)} \le \Delta\overline{D}_r$, and we have $|\Delta\overline{D}_f| \le \Delta\overline{D}_f$ and $|\Delta\overline{D}_r| \le \Delta\overline{D}_r$, which imply the suspension deflection performances can be achieved.

Remark 5. The works [38, 39] mainly focused on the finite-time control problem of quarter active suspension systems. Compared with the foregoing, this study not only addresses the finite-time control of half-vehicle but also further studies constraints of displacement and pitch angle of the car body. Particularly, the random faults between different actuators are studied.

4. Simulation Example

In this section, an example is given to demonstrate the established approach for half-vehicle ASSs. In order to facilitate the implement of the simulation, the nonlinear stiffness and dampers are given in [6]

$$\begin{aligned}
F_{\mathrm{sf}} &= k_{f01}\Delta D_f + k_{f02}\Delta D_f^3,\\
F_{\mathrm{sr}} &= k_{r01}\Delta D_r + k_{r02}\Delta D_r^3,\\
F_{\mathrm{tf}} &= k_{f2}\left(D_1 - D_{01}\right),\\
F_{\mathrm{tr}} &= k_{r2}\left(D_2 - D_{02}\right),\\
F_{\mathrm{bf}} &= b_{f2}\left(\dot{D}_1 - \dot{D}_{01}\right),\\
F_{\mathrm{br}} &= b_{r2}\left(\dot{D}_2 - \dot{D}_{02}\right),\\
F_{\mathrm{df}} &= \begin{cases} d_{e1}\Delta\dot{z}_f, \\ d_{c1}\Delta\dot{z}_f, \end{cases} F_{\mathrm{dr}} = \begin{cases} d_{e2}\Delta\dot{z}_r, \\ d_{c2}\Delta\dot{z}_r, \end{cases}
\end{aligned} \tag{67}$$

where k_{f01} and k_{f02} are the stiffness coefficients of linear and nonlinear terms for the front active suspension systems; k_{r01} and k_{r02} are the stiffness coefficients of linear and nonlinear terms for the rear active suspension systems; k_{f2} and b_{f2} stand for the damping coefficients of front and rear wheels; k_{r2} and b_{r2} denote the damping coefficients of front and rear wheels; d_{e1} and d_{c1} are the segmented coefficients of front damper of active suspension systems; and d_{e2} and d_{c2} are the

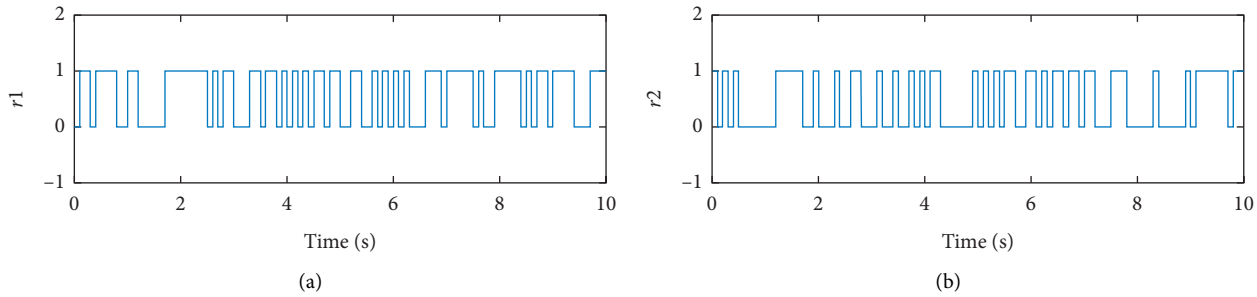


Figure 2: The curves of $r_1(t)$ and $r_2(t)$.

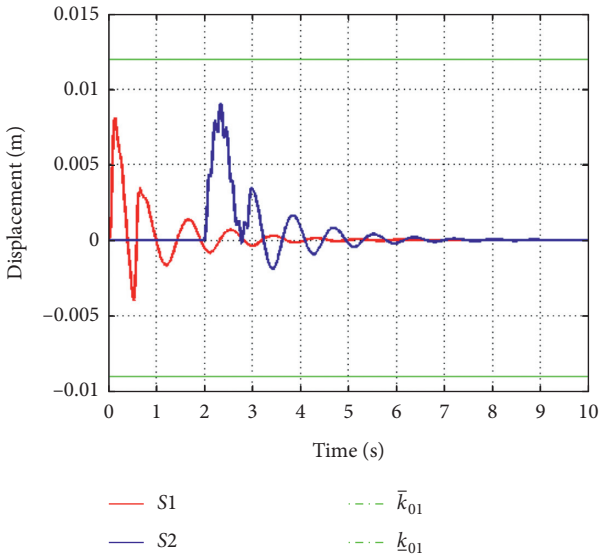


Figure 3: The vertical motion displacement x_1.

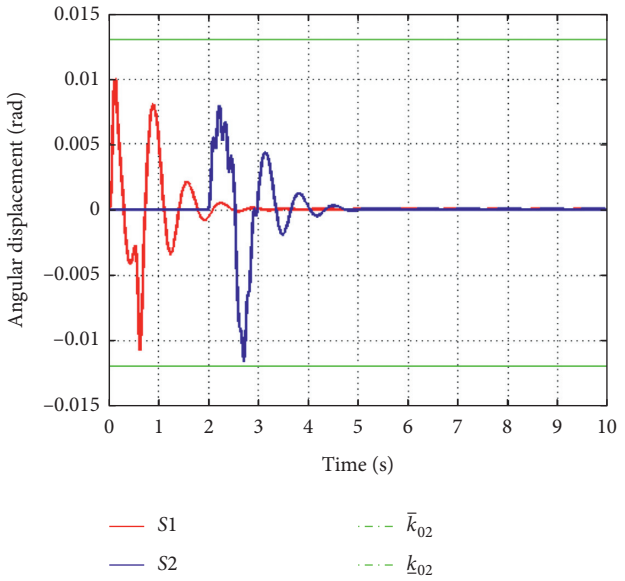


Figure 4: The pitch angle φ.

segmented coefficients of front damper of active suspension systems.

The road profile is an important aspect of affecting suspension performance. Thus, two different road inputs are carried out and given as follows:

*S*1: the road input of the front wheel is described as

$$D_{01} = \begin{cases} 0.5h\left[1 + \sin\left(\dfrac{2\pi v}{L}t\right)\right], & t_1 \le t \le t_h, \\ 0, & \text{otherwise}, \end{cases} \tag{68}$$

where h and L are the height and the length of the bump, and v is the velocity of the vehicle. The road input D_{02} for the rear wheel is implemented as $D_{02}(t) = D_{01}(t - \tau)$. Suppose $h = 0.05$ m, $L = 2.5$ m, $v = 18$ km/h, $t_1 = 2$, $t_h = 2.5$, and time delay $\tau = 0.5$.

*S*2: the road input of the front wheel is presented as

$$D_{01} = \begin{cases} 0.5h_b\left[1 + \sin\left(\dfrac{2\pi v_b}{L}t\right)\right], & t_1 \le t \le \dfrac{L}{v_b}, \\ 0, & t > \dfrac{L}{v_b}, \end{cases} \tag{69}$$

where $h_b = 0.08$ m, $L = 2.5$ m, and $v_b = 25$ km/h. The road input D_{02} is carried out the same as *S*1.

The parameters in half-vehicle active suspension systems are briefly illustrated: $M = 1200$ kg, $I = 1000$ kgm^2, $m_f = m_r = 100$ kg, $k_{f_{01}} = k_{r_{01}} = 15000$ N/m, $k_{f_{02}} = k_{r_{02}} = 1000$ N/m, $k_{f_2} = 200000$ N/m, $l_1 = 1.5$ m, $k_{r_2} = 150000$ N/m, $b_{f_2} = 1500$ Ns/m, $b_{r_2} = 2000$ Ns/m, $d_{e1} = d_{e2} = 1500$ Ns/m, $d_{c1} = d_{c2} = 1200$ Ns/m, and $l_2 = 1.2$ m. The control parameters in this work are given as follows: $\eta_1 = 4$, $\eta_2 = 6$, $\delta_1 = 3$, $\delta_2 = 2$, $\kappa_1 = 30$, $\kappa_2 = 40$, $c_1 = c_3 = 50$, $c_2 = c_4 = 100$, $a = 299/300$, $k_{a_1} = 0.006$,

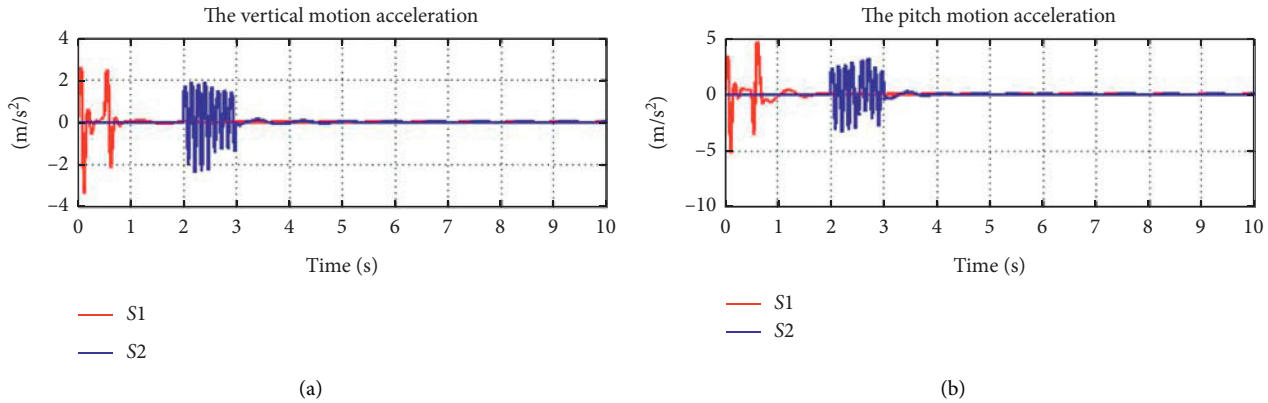


Figure 5: The accelerations of active suspension systems motion.

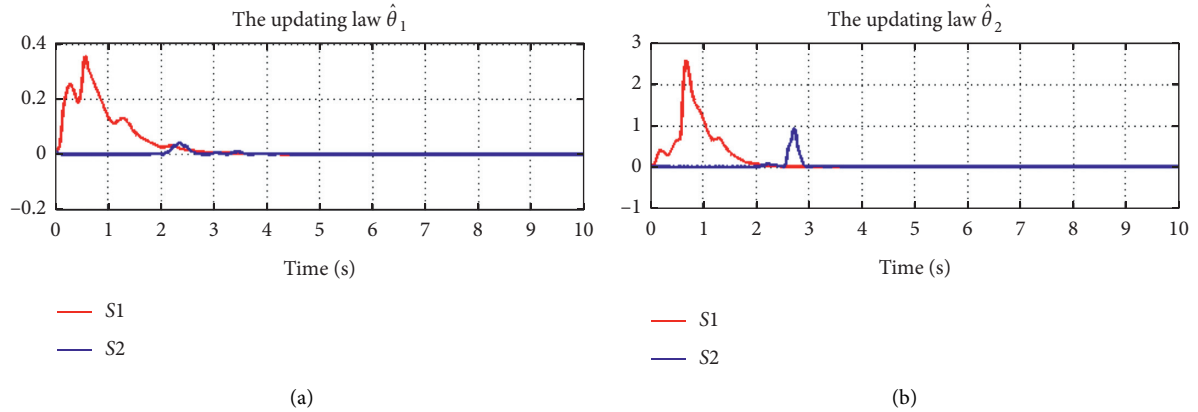


Figure 6: Trajectories of $\widehat{\theta}_i\,(i = 1, 2)$.

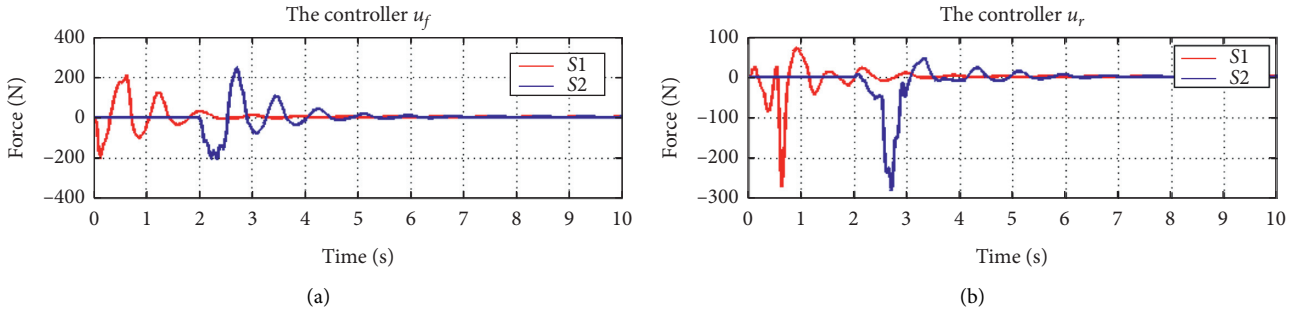


Figure 7: The controllers with fault tolerant.

$\underline{k}_{b_1} = 0.01$, $k_{a_2} = 0.012$, $k_{b_2} = 0.011$, $\overline{k}_{c_1} = 0.012$, $\underline{k}_{c_1} = 0.009$, $\overline{k}_{c_2} = 0.013$, $\underline{k}_{c_2} = 0.012$, and the rest ones are assumed as zeros. In simulation examples, the initial state values are given as $x(0) = [0, 0, 0, 0, 0, 0, 0, 0]^T$, $S\{0, 1\}$. For $r_1(t)$: $(\gamma_{pq})_{2\times 2} = \begin{bmatrix} -4 & 4 \\ 3 & -3 \end{bmatrix}$; $r_2(t)$: $(\gamma_{pq})_{2\times 2} = \begin{bmatrix} -3 & 3 \\ 4 & -4 \end{bmatrix}$.

For the cases $S1$ and $S2$, the curves of $r_1(t)$ and $r_2(t)$ are shown in Figure 2; the trajectories of x_1 which illustrates the displacement of vertical motion are shown in Figure 3. It can be seen that the displacement x_1 does not exceeds its constraint bounded.

The pitch angle φ is constrained under its bounded and can stabilize quickly in a finite time, as shown in Figure 4. Figure 5 shows the accelerations of vertical motion and pitch motion. It is observed that the maximums of vertical motion and pitch motion are $4\,\text{m/s}^2$ and $5\,\text{m/s}^2$, respectively. In addition, the stabilization of these motions is achieved in 1 sec for cases $S1$ and $S2$. The purpose of Figure 6 is to show the trajectories of $\widehat{\theta}_1$ and $\widehat{\theta}_2$. It indicates that for cases $S1$ and $S2$, they can converge to zero in a very short period of time. Figure 7 shows the controllers u_f and u_r, and we can observe that the control forces are calculated about 300 N. It means that the designed controllers play a key role in controlling

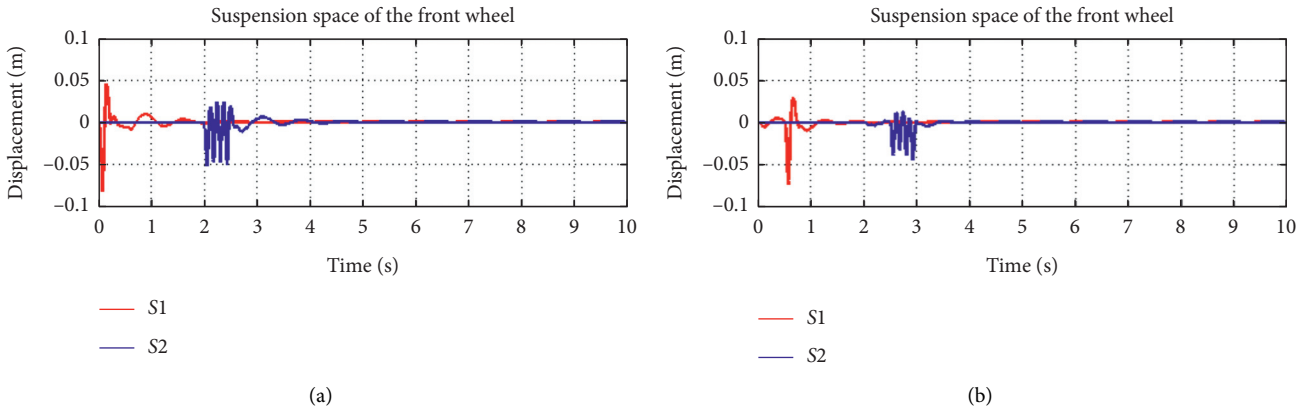


Figure 8: The suspension spaces.

the variation of active suspension systems. Figure 8 shows that the front wheel and rear wheel of the suspension spaces are all within the bounds. Through the above analysis, the proposed control schemes could be used to help reduce the variation of active suspension systems and achieve a good performance in a finite time.

5. Conclusions

In order to achieve good suspension performances, the finite-time control for half active suspension systems have been discussed in this study. The asymmetric BLFs have been employed to constrain vehicle body's displacement and pitch angle in a safety range. Then, utilizing the finite-time control design theory, the adaptive controllers were proposed for vertical motion and pitch motion, respectively. Finally, the effectiveness of designed schemes is demonstrated with the help of results of simulation examples. On the one hand, in the aspect of automobile research, we would further study the battery problem combined with the distributed controller of the multiagent consistency algorithm and apply it to distributed generators in the Energy Internet such as [56, 57]. On the other hand, we apply this method in the pipeline network system such as [58] and optimal deployment of agents in the pipeline network system in [59]. The study of control problem in the pipeline network system will be an interesting research topic.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (61473139 and 61903168), Provincial Key Research and Development Projects (JH2/10200002), and Provincial Scientific Research Fund of Education Department (JJL202015407).

References

[1] H. Pan, W. Sun, X. Jing, H. Gao, and J. Yao, "Adaptive tracking control for active suspension systems with non-ideal actuators," *Journal of Sound and Vibration*, vol. 399, pp. 2–20, 2017.

[2] W. Sun, H. Gao, and O. Kaynak, "Vibration isolation for active suspensions with performance constraints and actuator saturation," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 2, pp. 675–683, 2015.

[3] S. Kilicaslan, "Control of active suspension system considering nonlinear actuator dynamics," *Nonlinear Dynamics*, vol. 91, no. 2, pp. 1383–1394, 2018.

[4] W. Sun, H. Gao, and O. Kaynak, "Adaptive backstepping control for active suspension systems with hard constraints," *IEEE/ASME Transactions on Mechatronics*, vol. 18, no. 3, pp. 1072–1079, 2012.

[5] V. S. Deshpande, P. D. Shendge, and S. B. Phadke, "Nonlinear control for dual objective active suspension systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 656–665, 2016.

[6] F. Zhao, S. S. Ge, F. Tu, Y. Qin, and M. Dong, "Adaptive neural network control for active suspension system with actuator saturation," *IET Control Theory & Applications*, vol. 10, no. 14, pp. 1696–1705, 2016.

[7] G. Wang, C. Chen, and S. Yu, "Robust non-fragile finite-frequency H∞ static output-feedback control for active suspension systems," *Mechanical Systems and Signal Processing*, vol. 91, pp. 41–56, 2017.

[8] S.-A. Chen, J.-C. Wang, M. Yao, and Y.-B. Kim, "Improved optimal sliding mode control for a non-linear vehicle active suspension system," *Journal of Sound and Vibration*, vol. 395, pp. 1–25, 2017.

[9] C. Hua, J. Chen, Y. Li, and L. Li, "Adaptive prescribed performance control of half-car active suspension system with unknown dead-zone input," *Mechanical Systems and Signal Processing*, vol. 111, pp. 135–148, 2018.

[10] L. Liu, X. S. Li, Y. J. Liu, and S. C. Tong, "Neural network based adaptive event trigger control for a class of electromagnetic suspension systems," *Control Engineering Practice*, vol. 106, Article ID 104675, 2021.

[11] F. León-Vargas, F. Garelli, and M. Zapateiro, "Limiting vertical acceleration for ride comfort in active suspension systems," *Proceedings of the Institution of Mechanical*

Engineers, Part I: Journal of Systems and Control Engineering, vol. 232, no. 3, pp. 223–232, 2017.

[12] Y. Huang, J. Na, X. Wu, G.-B. Gao, and Y. Guo, "Robust adaptive control for vehicle active suspension systems with uncertain dynamics," *Transactions of the Institute of Measurement and Control*, vol. 40, no. 4, pp. 1237–1249, 2018.

[13] J. Na, Y. Huang, X. Wu, G. Gao, H. Guido, and Z. J. Jason, "Active adaptive estimation and control for vehicle suspensions with prescribed performance," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 6, pp. 2063–2077, 2017.

[14] W. Gao, N. Zhang, and H. P. Du, "A half-car model for dynamic analysis of vehicles with random parameters," *Proceedings of the 5th Australasian Congress on Applied Mechanics (ACAM 2007), Engineers Australia*, vol. 1, pp. 595–600, 2007.

[15] H. Li, J. Yu, C. Hilton, and H. Liu, "Adaptive sliding-mode control for nonlinear active suspension vehicle systems using T–S fuzzy approach," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 8, pp. 3328–3338, 2012.

[16] H. Li, H. Liu, C. Hilton, and S. Hand, "Non-fragile H∞ control for half-vehicle active suspension systems with actuator uncertainties," *Journal of Vibration and Control*, vol. 19, no. 4, pp. 560–575, 2013.

[17] S. Sui, C. L. P. Chen, and S. Tong, "A novel adaptive NN prescribed performance control for stochastic nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2020.

[18] D.-P. Li, D.-J. Li, Y.-J. Liu, S. Tong, and C. L. P. Chen, "Approximation-based adaptive neural tracking control of nonlinear MIMO unknown time-varying delay systems with full state constraints," *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3100–3109, 2017.

[19] W. He, Y. Dong, and C. Sun, "Adaptive neural network control of unknown nonlinear affine systems with input dead zone and output constraint," *ISA Transactions*, vol. 58, pp. 96–104, 2015.

[20] Z. Liu, G. Lai, Y. Zhang, and C. L. P. Chen, "Adaptive fuzzy tracking control of nonlinear time-delay systems with dead-zone output mechanism based on a novel smooth model," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 6, pp. 1998–2011, 2015.

[21] L. Liu, T. Gao, Y.-J. Liu, S. Tong, C. L. P. Chen, and L. Ma, "Time-varying IBLFs-based adaptive control of uncertain nonlinear systems with full state constraints," *Automatica*, vol. 129, Article ID 109595, 2021.

[22] W. He, Z. Yin, and C. Sun, "Adaptive neural network control of a marine vessel with constraints using the asymmetric barrier Lyapunov function," *IEEE Transactions on Cybernetics*, vol. 47, no. 7, pp. 1641–1651, 2017.

[23] Y. J. Liu, S. C. Tong, C. L. P. Chen, and D. J. Li, "Adaptive NN control using integral barrier Lyapunov functionals for uncertain nonlinear block-triangular constraint systems," *IEEE Transactions on Cybernetics*, vol. 47, no. 11, pp. 3747–3757, 2016.

[24] W. He, X. Mu, Y. Chen, X. He, and Y. Yu, "Modeling and vibration control of the flapping-wing robotic aircraft with output constraint," *Journal of Sound and Vibration*, vol. 423, pp. 472–483, 2018.

[25] W. Sun, H. Pan, Y. Zhang, and H. Gao, "Multi-objective control for uncertain nonlinear active suspension systems," *Mechatronics*, vol. 24, no. 4, pp. 318–327, 2013.

[26] Y.-J. Liu, Q. Zeng, S. Tong, C. L. P. Chen, and L. Liu, "Adaptive neural network control for active suspension systems with time-varying vertical displacement and speed constraints," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9458–9466, 2019.

[27] S. P. Bhat and D. S. Bernstein, "Continuous finite-time stabilization of the translational and rotational double integrators," *IEEE Transactions on Automatic Control*, vol. 43, no. 5, pp. 678–682, 2018.

[28] P. Zhao, D.-H. Zhai, Y. Sun, and Y. Li, "Adaptive finite-time control of a class of Markovian jump nonlinear systems with parametric and dynamic uncertainties," *Nonlinear Analysis: Hybrid Systems*, vol. 29, pp. 234–246, 2018.

[29] S. Sui, C. L. P. Chen, and S. Tong, "Event-trigger-based finite-time fuzzy adaptive control for stochastic nonlinear system with unmodeled dynamics," *IEEE Transactions on Fuzzy Systems*, vol. 29, 2020.

[30] R. Li and J. Cao, "Finite-time stability analysis for Markovian jump memristive neural networks with partly unknown transition probabilities," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 12, pp. 2924–2935, 2018.

[31] Y. Hong, J. Wang, and D. Cheng, "Adaptive finite-time control of nonlinear systems with parametric uncertainty," *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 858–862, 2005.

[32] Y. Dan, X. D. Li, and J. L. Qiu, "Output tracking control of delayed switched systems via state-dependent switching and dynamic output feedback," *Nonlinear Analysis: Hybrid Systems*, vol. 32, pp. 294–305, 2019.

[33] D. Yang, X. Li, J. Shen, and Z. Zhou, "State-dependent switching control of delayed switched systems with stable and unstable modes," *Mathematical Methods in the Applied Sciences*, vol. 41, no. 16, pp. 6968–6983, 2018.

[34] S. Sui, C. L. P. Chen, and S. Tong, "Neural-network-based adaptive DSC design for switched fractional-order nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2020.

[35] X. Zhang, G. Feng, and Y. Sun, "Finite-time stabilization by state feedback control for a class of time-varying nonlinear systems," *Automatica*, vol. 48, no. 3, pp. 499–504, 2012.

[36] H. Wang, B. Chen, C. Lin, Y. Sun, and F. Wang, "Adaptive finite-time control for a class of uncertain high-order non-linear systems based on fuzzy approximation," *IET Control Theory & Applications*, vol. 11, no. 5, pp. 677–684, 2017.

[37] M. Cai and Z. Xiang, "Adaptive practical finite-time stabilization for uncertain nonstrict feedback nonlinear systems with input nonlinearity," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 7, pp. 1668–1678, 2017.

[38] S. Sui, C. L. P. Chen, and S. C. Tong, "Fuzzy adaptive finite-time control design for nontriangular stochastic nonlinear systems," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 1, pp. 172–184, 2018.

[39] H. Pan, W. Sun, H. Gao, and J. Yu, "Finite-time stabilization for vehicle active suspension systems with hard constraints," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2663–2672, 2015.

[40] G. Wang, C. Chen, and S. Yu, "Finite-time sliding mode tracking control for active suspension systems via extended super-twisting observer," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 231, no. 6, pp. 459–470, 2017.

[41] Y. J. Liu, Y. Zhang, L. Liu, S. C. Tong, and C. L. P. Chen, "Adaptive finite-time control for half-vehicle active suspension systems with uncertain dynamics," *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 1, pp. 168–178, 2020.

[42] S. Zhao, H. Liang, P. Du, and S. Qi, "Adaptive NN finite-time tracking control of output constrained nonlinear system with input saturation," *Nonlinear Dynamics*, vol. 92, no. 4, pp. 1845–1856, 2018.

[43] R. Ma, Y. Liu, S. Zhao, and J. Fu, "Finite-time stabilization of a class of output-constrained nonlinear systems," *Journal of the Franklin Institute*, vol. 352, no. 12, pp. 5968–5984, 2015.

[44] Z. Gao and P. Antsaklis, "Stability of the pseudo-inverse method for reconfigurable control systems," *International Journal of Control*, vol. 53, no. 3, pp. 717–729, 1996.

[45] M. M. Kale and A. J. Chipperfield, "Stabilized MPC formulations for robust reconfigurable flight control," *Control Engineering Practice*, vol. 13, no. 6, pp. 771–788, 2005.

[46] M. L. Corradini and G. Orlando, "Actuator failure identification and compensation through sliding modes," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 1, pp. 184–190, 2006.

[47] G. Tao, S. M. Joshi, and X. Ma, "Adaptive state feedback and tracking control of systems with actuator failures," *IEEE Transactions on Automatic Control*, vol. 46, no. 1, pp. 78–95, 2001.

[48] G. Lai, C. Wen, Z. Liu, Y. Zhang, and S. Xie, "Adaptive compensation for infinite number of actuator failures/faults using output feedback control," *Information Sciences*, vol. 399, pp. 1–12, 2017.

[49] L. Liu, Y.-J. Liu, and S. Tong, "Neural networks-based adaptive finite-time fault-tolerant control for a class of strict-feedback switched nonlinear systems," *IEEE Transactions on Cybernetics*, vol. 49, no. 7, pp. 2536–2545, 2019.

[50] Y.-J. Liu, S. Li, S. Tong, and C. L. P. Chen, "Neural approximation-based adaptive control for a class of nonlinear nonstrict feedback discrete-time systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 7, pp. 1531–1541, 2017.

[51] O. L. V. Costa, M. D. Fragoso, and R. P. Marques, *Discrete Time Markov Jump Linear Systems*, Springer-Verlag, New York, NY, USA, 2005.

[52] H. Fan, B. Liu, Y. Shen, and W. Wang, "Adaptive failure compensation control for uncertain systems with stochastic actuator failures," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 808–814, 2014.

[53] B. Liu, M. Saif, and H. Fan, "Adaptive fault tolerant control of a half-car active suspension systems subject to random actuator failures," *IEEE/ASME Transactions on Mechatronics*, vol. 21, no. 6, pp. 2847–2857, 2016.

[54] J. Yin, S. Khoo, Z. Man, and X. Yu, "Finite-time stability and instability of stochastic nonlinear systems," *Automatica*, vol. 47, no. 12, pp. 2671–2677, 2011.

[55] S. Khoo, J. Yin, Z. Man, and X. Yu, "Finite-time stabilization of stochastic nonlinear systems in strict-feedback form," *Automatica*, vol. 49, no. 5, pp. 1403–1410, 2013.

[56] R. Wang, Q. Sun, W. Hu, Y. Li, D. Ma, and P. Wang, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, pp. 7866–7879, 2021.

[57] Q. Sun, R. Han, H. Zhang, J. Zhou, and J. M. Guerrero, "A multiagent-based consensus algorithm for distributed coordinated control of distributed generators in the energy internet," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 3006–3019, 2015.

[58] X. Hu, H. Zhang, D. Ma, and R. Wang, "Hierarchical pressure data recovery for pipeline network via generative adversarial networks," *IEEE Transactions on Automation Science and Engineering*, pp. 1–11, 2021.

[59] D. Ma, Y. Li, X. Hu, H. Zhang, and X. Xie, "An optimal three-dimensional drone layout method for maximum signal coverage and minimum interference in complex pipeline networks," *IEEE Transactions on Cybernetics*, pp. 1–11, 2021.

Hindawi
Complexity
Volume 2021, Article ID 5549776, 16 pages
https://doi.org/10.1155/2021/5549776

Research Article

Trajectory Tracking Control in Real-Time of Dual-Motor-Driven Driverless Racing Car Based on Optimal Control Theory and Fuzzy Logic Method

Gang Li,[1,2] **Sucai Zhang**,[1] **Lei Liu**,[3] **Xubin Zhang**,[1] **and Yuming Yin**[4]

[1]*School of Automobile and Traffic Engineering, Liaoning University of Technology, Jinzhou 121001, Liaoning, China*
[2]*Beijing University of Science and Technology Information, Beijing, China*
[3]*College of Science, Liaoning University of Technology, Jinzhou 121001, Liaoning, China*
[4]*School of Vehicle and Transportation, Tsinghua University, Beijing, China*

Correspondence should be addressed to Gang Li; qcxyligang@lnut.edu.cn

Received 22 February 2021; Revised 17 March 2021; Accepted 27 March 2021; Published 29 April 2021

Academic Editor: Rui Wang

To improve the accuracy and timeliness of the trajectory tracking control of the driverless racing car during the race, this paper proposes a track tracking control method that integrates the rear wheel differential drive and the front wheel active steering based on optimal control theory and fuzzy logic method. The model of the lateral track tracking error of the racing car is established. The model is linearized and discretized, and the quadratic optimal steering control problem is constructed. Taking advantage of the differential drive of dual-motor-driven racing car, the dual motors differential drive fuzzy controller is designed and integrated driving with active steering control. Simulation analysis and actual car verification show that this integrated control method can ensure that the car tracks different race tracks well and improve the track tracking control accuracy by nearly 30%.

1. Introduction

The Formula Student Autonomous China has been successfully held for three years and provides an important practical platform for cultivating talents in China's new energy vehicle and driverless industry. Trajectory tracking control is one of the key technologies of driverless car. A lot of related work has been carried out at home and abroad in the research of trajectory tracking and control of driverless car. Abroad, the University of Upper Alsace in France carried out the horizontal and vertical decoupling control of autonomous vehicles. Based on the theory of nonlinear model predictive control, the vehicle control strategy was designed to improve the accuracy and stability of vehicle trajectory tracking control [1]. The University of São Paulo in Brazil proposed a front wheel steering controller based on a dynamic inverse model for front wheel drive vehicles, which can make the car track the reference path under the tire adhesion limit and respond well to the discontinuity of the path curvature [2]. Stanford University in the United States designed a feedback-feedforward steering controller, which not only meets the minimum tracking path deviation, but also ensures the steering stability of the car [3]. The MPC method is used to realize the local trajectory planning and trajectory tracking during the collision avoidance process of the autonomous driving vehicle, which ensures the stability and safety of the autonomous vehicle during the collision avoidance process [4]. The Ford Motor Technology Research Center in the United States designed a linear state feedback controller and used the electric power steering system to realize the avoidance path tracking control of autonomous vehicles [5]. The Tokyo University of Agriculture and Technology in Japan combines the haptic steering torque guidance system and the direct yaw torque control strategy, and the simulation experiment verifies that the algorithm can well improve the vehicle trajectory tracking ability [6]. The University of California proposed a "dual loop" control method. The outer loop uses a nonlinear nonconvex model

predictive control method to plan the collision avoidance path, and the inner loop uses linear feedback control to track the desired trajectory. This method's effectiveness was verified by simulation experiments [7]. The University of Sheffield in the United Kingdom proposed a time-varying optimal controller to solve the problem of uncertain effects of dynamic model parameters and control parameters in the trajectory tracking control process [8].

In China, Tsinghua University proposed a path tracking method based on BP neural network PID controller [9]. An adaptive neural network control method is proposed, which realizes the robust control under the condition of uncertain tire cornering stiffness and unknown disturbance of the system in emergency avoidance conditions of autonomous vehicles [10]. Jilin University has designed a vehicle trajectory tracking controller based on linear model predictive control theory, which solves the trajectory tracking problem of double safety envelopes. Simulation experiments verify that the control method can effectively improve the trajectory tracking accuracy [11]. Beijing Institute of Technology proposes a safety envelope control strategy to ensure that autonomous vehicles can track the expect trajectory effectively [12]. The model predictive control theory is used to design the stability control strategy of the autonomous vehicle during high-speed driving considering the terrain factors of the vehicle [13]. A fuzzy controller is designed to control the preview distance and reduce the trajectory tracking deviation from Hefei University of Technology and verified the effectiveness of the control method through actual car experiments [14]. Dalian University of Technology builds a closed-loop control structure for vehicle trajectory tracking through integral inversion theory and designs a trajectory tracking controller to ensure real-time trajectory tracking and position deviation convergence [15]. Lei Liu et al. proposed an adaptive control method for a class of uncertain nonlinear systems with strict feedback. Finally, a numerical simulation is carried out to further verify the importance and feasibility of the proposed control scheme [16]. The above research methods provide the good references for the trajectory tracking and control of driverless race car.

However, the above research mainly focuses on traditional front wheel drive vehicles and realizes trajectory tracking control by controlling the active steering of the front wheels, and the algorithm based on model prediction has high requirements for the calculation ability of the vehicle controller, which is difficult to apply to the actual car verification. While the dual-motor-driven driverless car is driving on a specific track, it is more curved and complex than other driverless vehicles. Therefore, under this working condition, the racing car takes advantage of the differential drive of dual-motor-driven driverless car, which can significantly improve the tracking accuracy of the driverless racing car.

In this paper, the trajectory tracking of the driverless racing car driven by dual motors on the rear axle independently under track conditions is to ensure that the car has better track tracking accuracy in the track conditions and high real-time tracking. A method of trajectory tracking control is proposed. Firstly, in view of the large amount of calculation of the lateral control model predictive control algorithm, it is proposed to use the LQR control algorithm to improve the real-time performance of tracking. Secondly, in order to improve the accuracy of trajectory tracking, a method of integrating optimal steering control and differential drive control is proposed. Finally, the algorithm is verified through software simulation and actual car test.

2. Tracking Control Strategy of Driverless Racing Car

The integrated control strategy of dual-motor-driven driverless racing car track tracking is shown in Figure 1. It mainly includes the combination of feedforward-feedback steering control and dual-motor differential drive control. The steering control part adopts the optimal control algorithm. Through the lateral tracking error state equation of the driverless racing car, the lateral deviation, lateral deviation change, heading deviation, and heading deviation change between the actual state and the desired tracking trajectory are used as the input and output of the optimal controller. The control amount is the steering wheel angle. The dual-motor-driven part uses the error of the yaw rate to calculate the additional yaw moment through the fuzzy controller based on the fuzzy method, which obtains the differential drive torque according to the torque conversion principle and then obtains the drive torque of the left and right drive wheels based on the torque distribution method. This integrated control method can improve the accuracy and stability of the driverless racing car in the trajectory tracking process.

3. Optimal Steering Control Algorithm

This paper uses the vehicle two-degree-of-freedom model to study the lateral and yaw motion of the vehicle, as shown in Figure 2. Two dynamic degrees of freedom include the lateral position y of the car and the racing angle φ of the car. The lateral position y of the car is the distance from the car to the center of rotation. The racing direction angle φ is between the X axis of the Earth coordinate system and the front of the car. The longitudinal speed at the center of mass of the vehicle is represented by V_x.

Ignoring the influence of road slope, along the y-axis:

$$ma_y = F_{yf} + F_{yr}. \tag{1}$$

Lateral acceleration a_y is composed of movement acceleration $\ddot{y}$ along the y-axis and centripetal acceleration $v_y\dot{\varphi}$:

$$a_y = \ddot{y} + v_x\dot{\varphi}. \tag{2}$$

Combine the two types to get

$$m\left(\ddot{y} + v_x\dot{\varphi}\right) = F_{yf} + F_{yr}. \tag{3}$$

Find the balance equation by the torque around the z-axis, and finally get the vehicle yaw dynamic equation:

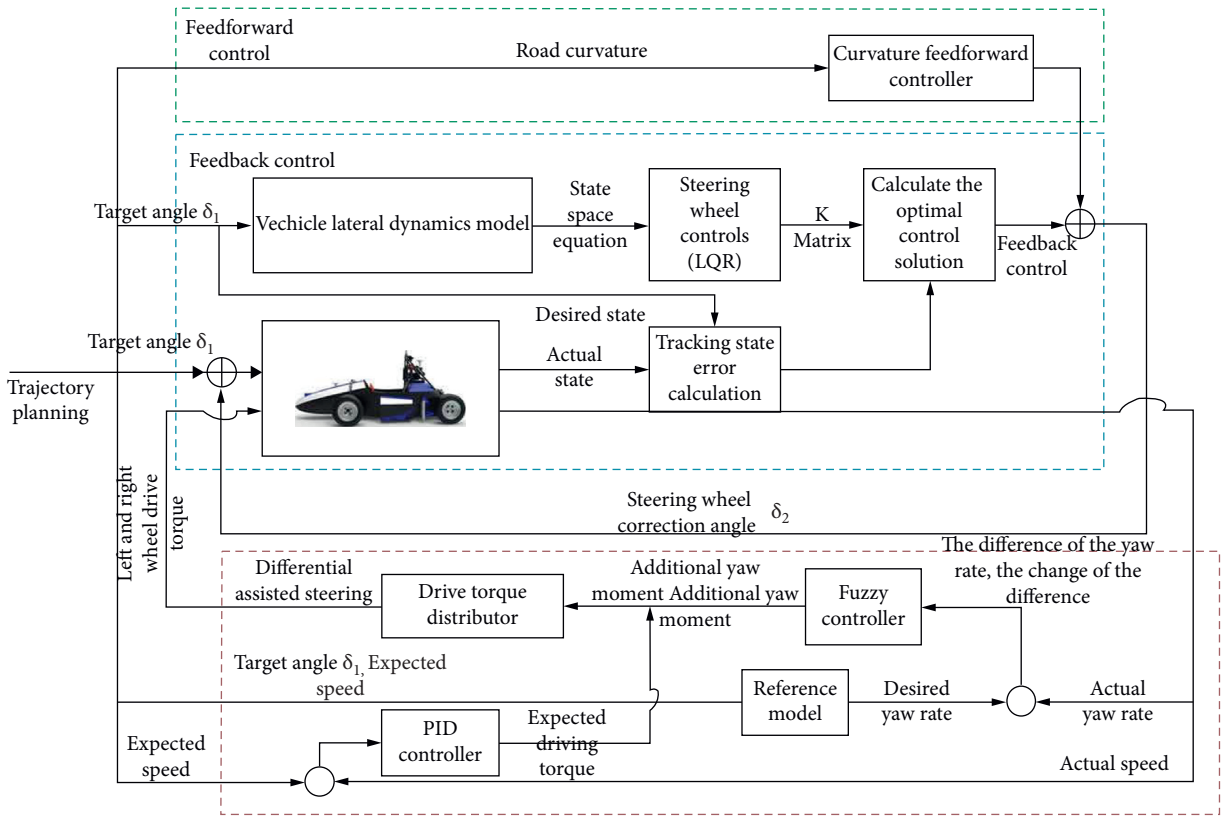


Figure 1: Integrated control strategy for trajectory tracking of driverless car.

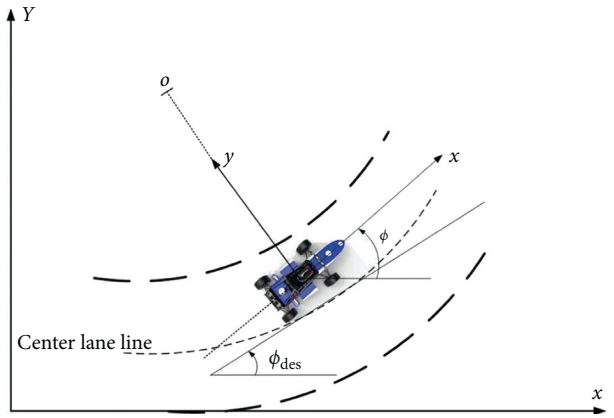


Figure 2: Lateral dynamics model of driverless car.

$$I_z\ddot{\varphi} = aF_{yf} - bF_{yr}. \tag{4}$$

According to the tire dynamics characteristics, when the side slip angle is small, the side force of the tire is proportional to the "side slip angle." From Figure 3, the side slip angle of the driverless racing car tire can be obtained. The front wheel slip angle:

$$\alpha_f = \delta - \delta_f. \tag{5}$$

From this, the vehicle lateral tracking error state equation is built. The rear wheel slip angle can be approximately expressed as

$$\alpha_r = -\delta_r. \tag{6}$$

Therefore, the lateral force of the front and rear wheels of the driverless racing car can be expressed as

$$\begin{cases} F_{yf} = 2C_{\alpha f}\left(\delta - \delta_f\right), \\ F_{yr} = 2C_{\alpha r}\left(-\delta_r\right). \end{cases} \tag{7}$$

From the geometric characteristics of kinematics:

$$\begin{cases} \tan\left(\delta_f\right) = \dfrac{v_y + a\dot{\varphi}}{v_x}, \\ \tan\left(\delta_r\right) = \dfrac{v_y - b\dot{\varphi}}{v_x}. \end{cases} \tag{8}$$

When the slip angle is small, the approximate angle and abbreviation $v_y = \dot{y}$, and (8) can be abbreviated to as (9).

$$\begin{cases} \delta_f = \dfrac{\dot{y} + a\dot{\varphi}}{v_x}, \\ \delta_r = \dfrac{\dot{y} - b\dot{\varphi}}{v_x}. \end{cases} \tag{9}$$

Finally, the state equation model of the driverless car is shown in (10).

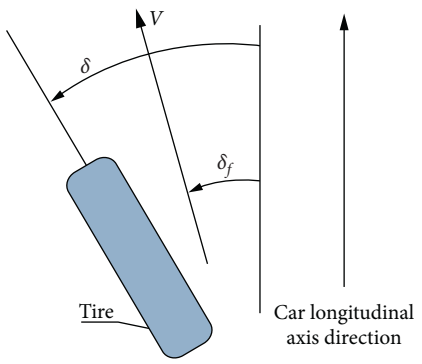


Figure 3: Side slip angle of driverless car tires.

Table 1: Parameters of driverless car.

Parameters' name	Parameters' value
Wheelbase L	1550 mm
Distance from center of mass to front axle of car a	1220 mm
Distance from center of mass to rear axle b	1200 mm
Vehicle quality m	310 kg
Front wheel cornering stiffness $C_{\alpha f}$	95202
Rear wheel cornering stiffness $C_{\alpha r}$	63947
Length/Width/Height	2860/1120/1175 mm
Front/Rear axle track	1220/1200 mm
Centroid height	280 mm
Moment of inertiaI_z around Z-axis	340 Kg m^2
Tire model	18.0 × 7.5–10R25 B
Wheel radius	237 mm
Curb quality	310 Kg

Table 2: Definition of symbols in the lateral tracking model of driverless racing car.

a_x, a_y	Components of acceleration on the x- and y-axes
v_x, v_y	Racing horizontal and vertical speed
φ	Heading angle
δ_f, δ_r	Speed direction of front and rear wheels
α_f, α_r	Side slip angle of front and rear wheels
δ	Front wheel steering angle
R	Turning radius

Table 3: Fuzzy control rule table.

Rules		Deviation				
	Torque	NB	NS	ZE	PS	PB
Deviation change	NB	PB	PM	PM	PS	ZE
	NS	PB	PM	PS	ZE	NM
	ZE	PB	PM	ZE	NM	NB
	PS	PM	ZE	NS	NM	NB
	PB	ZE	NS	NM	NB	NB

$$\frac{d}{dt}\begin{Bmatrix} y \\ \dot{y} \\ \varphi \\ \dot{\varphi} \end{Bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -\dfrac{2C_{\alpha f}+2C_{\alpha r}}{mv_x} & 0 & -v_x-\dfrac{2aC_{\alpha f}-2bC_{\alpha r}}{mv_x} \\ 0 & 0 & 0 & 1 \\ 0 & -\dfrac{2aC_{\alpha f}-2bC_{\alpha r}}{I_z v_x} & 0 & -\dfrac{2a^2C_{\alpha f}-2b^2C_{\alpha r}}{I_z v_x} \end{bmatrix} + \delta * \begin{bmatrix} 0 \\ \dfrac{2C_{\alpha f}}{m} \\ 0 \\ \dfrac{2aC_{\alpha f}}{I_z} \end{bmatrix}. \tag{10}$$

The racing car parameters are shown in Table 1, and Table 2 shows the definition of symbols used in the above expressions.

The lateral control model of the driverless racing car obtained above is mainly realized by controlling the tire rotation angle. For the driver, the steering wheel angle of the racing car can be directly controlled. The tracking error of the desired path direction and distance is the system variable to build a racing dynamics model. Let e_l be the lateral error; that is, it stands for the distance between the center of mass of the car and the desired tracking trajectory, and e_φ is the heading error; that is, the difference between the actual heading angle of the car and the desired heading angle. It can be seen that the speed required for the car to turn the desired angle is

$$\dot{\varphi}_{\text{des}} = \frac{v_x}{R}. \tag{11}$$

The lateral acceleration required for racing is

$$a_{\text{des}} = \frac{v_x^2}{R} = v_x \dot{\varphi}_{\text{des}}. \tag{12}$$

Therefore, the lateral acceleration error of the car is

$$\ddot{e}_l = a_y - a_{\text{des}} = \left(\ddot{y} + v_x\dot{\varphi}\right) - \frac{v_x^2}{R} = \ddot{y} + v_x\left(\dot{\varphi} - \dot{\varphi}_{\text{des}}\right). \quad (13)$$

The lateral speed error of the car is

$$\ddot{e}_l = \dot{y} + v_x\left(\varphi - \varphi_{\text{des}}\right). \quad (14)$$

Then the heading error of the car is shown in (15):

$$e_\varphi = \varphi - \varphi_{\text{des}}. \quad (15)$$

Combining and simplifying the above formulas (10) to (15), (16) can be obtained:

$$\dot{x} = Ax + B_1\delta + B_2\omega_{\text{des}}, \quad (16)$$

where ω_{des} is the yaw rate in the desired trajectory, $x = [e_l \ \dot{e}_l \ e_\theta \ \dot{e}_\theta]$. The matrices A, B_1, and B_2 are

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -(2C_{\alpha f} + 2C_{\alpha r}/mv_x) & (2C_{\alpha f} + 2C_{\alpha r}/m) & (-2aC_{\alpha f} + 2bC_{\alpha r}/mv_x) \\ 0 & 0 & 0 & 1 \\ 0 & -(2aC_{\alpha f} - 2bC_{\alpha r}/I_z v_x) & (2aC_{\alpha f} - 2bC_{\alpha r}/I_z) & -(2a^2C_{\alpha f} + 2b^2C_{\alpha r}/I_z v_x) \end{bmatrix}$$

$$B_1 = \begin{bmatrix} 0 \\ (2C_{\alpha f}/m) \\ 0 \\ (2C_{\alpha fa}/I_z) \end{bmatrix}.$$

Since driverless cars are highly complex nonlinear time-varying systems using linear quadratic optimal control methods, the above models must first be linearized and discretized. This paper uses the bilinear transformation discrete method to obtain the state matrix in discrete time, and the discrete time is set to $T = 0.01$s.

$$\begin{aligned} A_t &= \left(\frac{E + T/2 * A}{E - T/2 * A}\right), \\ B_{1t} &= B_1 * T, \\ B_{2t} &= B_2 * T. \end{aligned} \quad (17)$$

Since the lateral control is mainly for the control of the lateral deviation, the rate of change of the lateral deviation, the heading deviation, and the rate of change of the heading deviation, the term $B_2\dot{\varphi}_{qw}$ in the formula is ignored. The cost function J of steering control is

$$J = \sum_{t=0}^{N-1}\left(x_t^T Q x_t + \delta_t^T R\delta_t\right) + x_N^T Q_f x_N, \quad (18)$$

where Q is the state weighting matrix, R is the control variable weighting matrix, Q_f is the final state weighting matrix, x_t is the tracking error state variable, δ_t is the control variable, and x_N is the final state tracking error.

The optimal front wheel feedback control angle form for designing a racing car is

$$\delta^*(t) = -K(t)x(t). \quad (19)$$

Carrying out the convex optimization construction and solving the Riccati equation to (8), the optimal feedback control law can be obtained as

$$K(t) = -\left(R(t) + B_{1t}^T P_t B_{1t}\right)^{-1} B_{1t}^T P A_t. \quad (20)$$

R (t) is combined with feedforward steering control. The expected steering wheel angle is

$$\delta = \delta(t) + \delta^*(t). \quad (21)$$

In this formula, $\delta^*(t)$ is the feedforward control amount of the steering wheel angle.

The weighting matrix selected in this article is

$Q = [300\ 0\ 0\ 0;\ 0\ 50\ 0\ 0;\ 0\ 0\ 200\ 0;\ 0\ 0\ 0\ 50]$, $R = [10]$, $Q_N = [1\ 0\ 0\ 0;\ 0\ 1\ 0\ 0;\ 0\ 0\ 1\ 0;\ 0\ 0\ 0\ 0]$.

This section adopts the full-state feedback control of driverless racing car, including the lateral position deviation, lateral position deviation change rate, heading angle deviation, and heading angle deviation change rate during the tracking process. The calculation of the lateral error is shown in Figure 4.

The racing tracking state error model based on the right-hand rule is built in the lateral control of the trajectory tracking of the driverless racing car, so the racing tracking lateral error is composed of the difference between the actual running state of the racing car and the expected trajectory curve. The specific expression is shown as (24).

$$\begin{cases} e_l = dy\ \cos\left(\varphi_{\text{des}}\right) - dx\ \sin\left(\varphi_{\text{des}}\right), \\ \dot{e}_l = v_x \sin\left(\Delta\varphi\right) = v_x \sin\left(e_\varphi\right), \\ e_\varphi = \varphi - \varphi_{\text{des}}, \\ \dot{e}_\varphi = \dot{\varphi} - \dot{\varphi}_{\text{des}}. \end{cases} \quad (22)$$

Among them, $\dot{\varphi}$ is the vehicle yaw rate, which can be obtained by the vehicle body yaw rate sensor on the racing car, and $\dot{\varphi}_{\text{des}}$ is the expected yaw rate of the racing car, which is obtained by the desired path line fitting and desired tracking trajectory point solution controller.

4. Differential Drive Control Algorithm

The yaw moment of the racing car can increase the side slip angle of the driving wheel and reduce the side slip angle of the steering wheel, thereby indirectly reducing the steering radius of the racing car during high-speed driving, and has the function of assisting steering. In this paper, a fuzzy control method is adopted. The difference between the actual yaw rate and the desired yaw rate and the rate of change of the difference are used as the input of the fuzzy controller to output the additional yaw moment of the car, which is then converted into the differential driving moment of the driving wheels. Under the premise that the total driving torque remains unchanged, the left and right driving wheels are reduced or increased by 1/2 of the driving torque to realize the redistribution of the driving torque of the driving wheels, as shown in Figure 5.

Based on the two-degree-of-freedom vehicle model, the desired yaw rate of the car can be calculated as

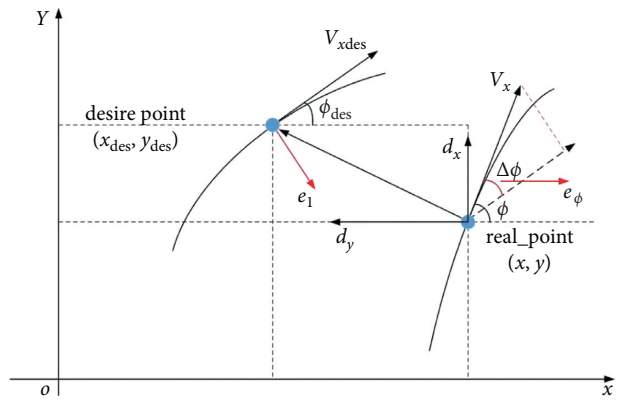


Figure 4: Schematic diagram of calculation of lateral error of driverless car.

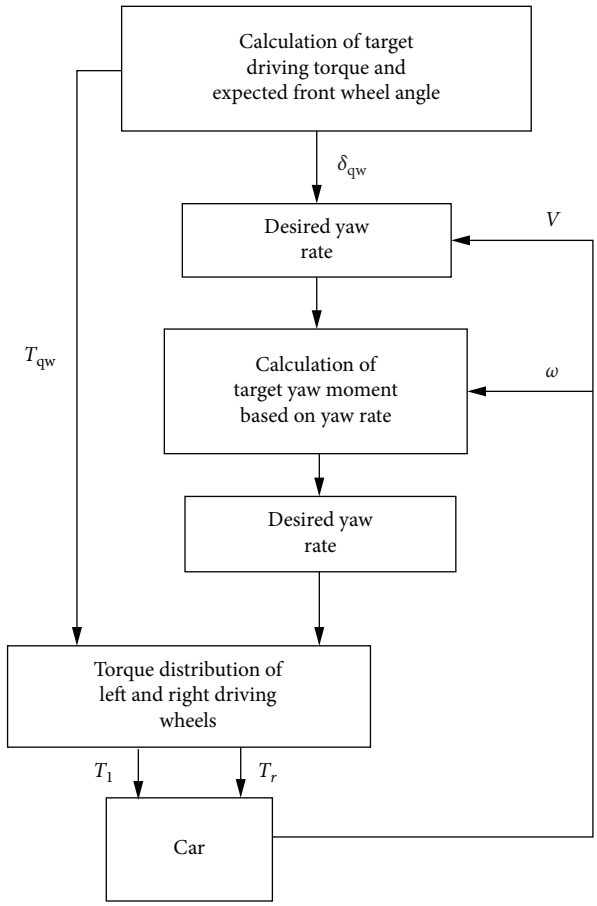


Figure 5: The driving torque distribution strategy based on the desired yaw rate.

$$\omega_{qw} = \frac{v_x/L}{1 + Kv_x^2}\delta,$$
$$K = \frac{m}{L^2}\left(\frac{b}{C_r} - \frac{a}{C_f}\right). \tag{23}$$

In the formula, K is the steering stability factor of the car.

Set the ideal side slip angle of the center of mass, namely, $X = 0$, so that the vehicle can track the desired trajectory as much as possible. When calculating the yaw rate, taking into account the limit adhesion coefficient of the road, the range of the car's yaw rate is

$$|\omega| \le \frac{\mu_{\max} g}{v_x}. \tag{24}$$

Then the desired yaw rate is

$$\omega_{qw} = \min\left\{\left|\frac{v_x/L}{1 + Kv_x^2}\delta\right|, \left|\frac{\mu_{\max} g}{v_x}\right|\right\}\operatorname{sgn}(\delta). \tag{25}$$

In the formula, $\mu_{\max}$ is the peak adhesion coefficient between the road and racing tires.

This paper uses the yaw rate error and the rate of change of the error as the input of the fuzzy controller, and the output is the additional yaw moment. Three membership functions are formed by using common triangular membership functions and trapezoidal membership functions. The yaw rate deviation and the rate of change of the deviation are taken as (−1, 1). The range of change is [−3, −2, −1, 0, 1, 2, 3]. Each defines 5 fuzzy subsets, namely, NB (large negative), NS (small negative), ZE (zero), PS (positive small), and PB (positive large). Figure 6 is the membership function of the yaw rate error, and Figure 7 is the membership function of the yaw rate change rate.

The fuzzy control rule table is shown Table 3. According to the fuzzy rules formulated by experts in accordance with the theory of racing dynamics, it includes the control strategy of the controlled yaw moment under different working conditions, which can significantly assist the steering of the car and reduce the trajectory tracking error.

The fuzzy subsets obtained by fuzzy inference based on the fuzzy rule library need to be converted into accurate values. Two clarification methods are commonly used: the maximum membership method and the center of gravity method. The area center of gravity method is used to defuzzify to obtain the accurate value of the inference output. In actual control, the defuzzy output needs to be multiplied by the corresponding scale factor to obtain the final yaw moment for control.

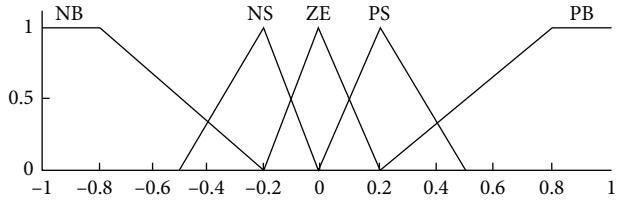


Figure 6: Membership function of yaw rate error.

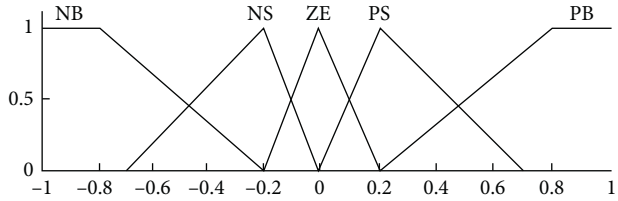


Figure 7: Membership function of yaw rate change rate.

The designed fuzzy controller has two inputs and one output. The difference between the actual yaw rate and the desired yaw rate and the change of the difference are used as the input. The output is the additional yaw moment M_{fj}, and at this moment the differential torque is obtained according to the dual motors independently driving the differential torque, and the method of using the equal division method to increase or decrease the left and right driving wheels by the same driving torque ΔT:

$$M_{fj} = \Delta T r_{\text{tire}} \frac{h}{2} + \Delta T r_{\text{tire}} \frac{h}{2} = \Delta T r_{\text{tire}} h,$$
$$\Delta T = \frac{M_{fj}}{r_{\text{tire}} h}. \tag{26}$$

In the formula, r_{tire} is the rolling radius of the driving wheel.

In order to ensure that the total expected driving torque of the car remains unchanged, the same torque distribution rule is adopted to rematch the driving torque of the left and right driving wheels, which are

$$\begin{cases} T_l = T_{qw} - \Delta T, \\ T_r = T_{qw} + \Delta T. \end{cases} \tag{27}$$

5. Simulation and Experimental Verification Results

Firstly, CarSim and Matlab/Simulink cosimulation was used to simulate and verify the algorithm, and then the actual car was verified on the dual-motor-driven driverless racing car. Therein, the detail simulator is similar to the literatures [17, 18]. Since the width of the track during the race is 3.5 m, and there are many corners and rapid changes in curvature, the low speed of the car in this working condition ensures that it can drive stably. Set two working conditions to verify the algorithm and they are straight-line trajectory tracking and curved trajectory tracking.

5.1. Straight-Line Tracking. The trajectory controller is built in CarSim and Matlab/Simulink, which tracks the desired path. Set the vehicle speed to 3 m/s, and track the desired trajectory under straight-line working conditions. The simulation results of straight-line working conditions are shown in Figures 7–10. The specific analysis is as follows.

It can be seen from Figure 8 that although the initial position is set to a lateral deviation of 0.5 m, the car can track the desired trajectory when driving 3.5 m. At the same time, the tracking error of the longitudinal vehicle speed gradually tends to zero after a short fluctuation, and it is shown as Figure 9. From Figure 10, it can be seen that, during this process, the front wheel steering angle fluctuates relatively a little, and it is basically stable at 2s, which meets the requirements of transient.

5.2. Curve Path Tracking. In curve tracking, the minimum turning radius is set to 3.5 m. The car still tracks the desired trajectory at the speed of 3 m/s. The comparison diagram between the actual tracking trajectory and the expected trajectory curve is shown in Figure 11.

Although the curve of the track is very sharp, it can be seen that the error value is less than 0.2 m from Figure 12, which can track the expected trajectory well. From Figure 13, it can be seen that the total steering is composed of the target rotation angle and the feedback rotation angle. The angle changes smoothly. When there is a lateral tracking error, the feedback corner will correct the target corner, which is the expected trajectory of the racing car. It can be seen from Figure 14 that the racing car tracks the expected speed in 3s, and then a slight fluctuation is added to the expected speed.

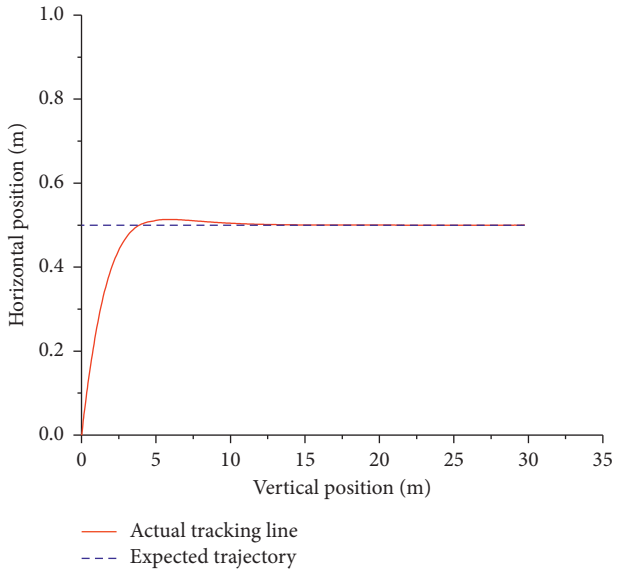


Figure 8: Position change during linear trajectory tracking.

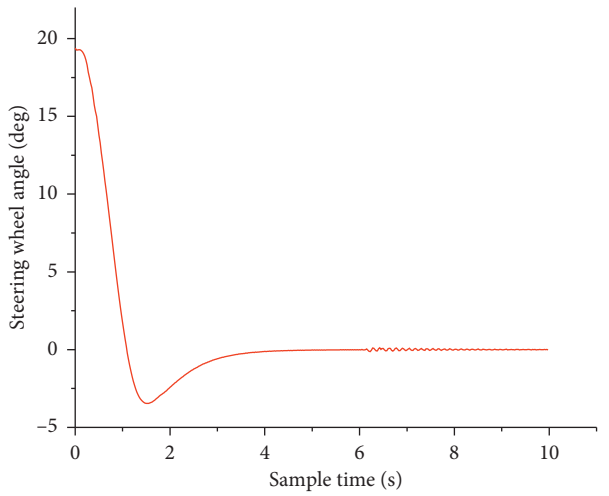


Figure 10: Simulation diagram of straight-line tracking steering angle.

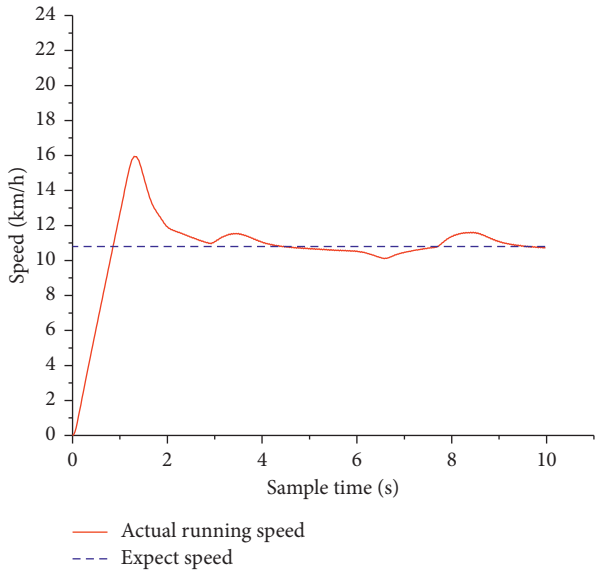


Figure 9: Straight-line tracking vehicle speed simulation diagram.

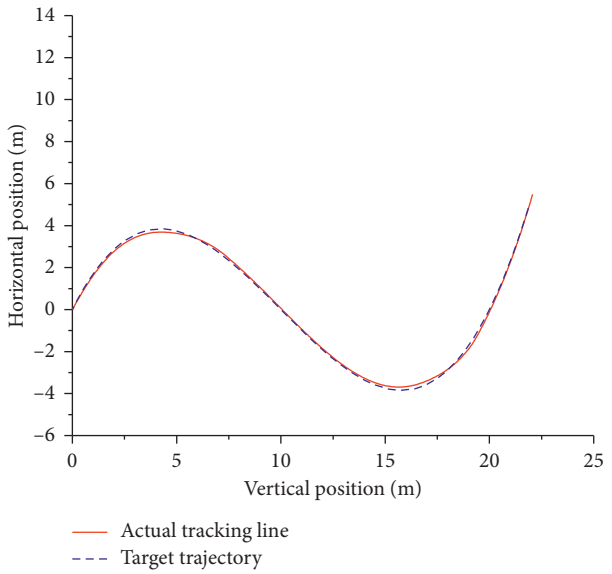


Figure 11: Path position tracking simulation diagram.

5.3. Integrated Control Simulation Verification. Establish a CarSim/Simulink cosimulation control model. The current vehicle speed, total steering wheel angle, and expected drive torque feedback by the racing car are input into the drive torque distribution controller based on the yaw rate, and the drive torque obtained by the left and right driving wheels is calculated. The rotational speed of the driving wheel and the obtained driving torque are input to the corresponding driving motor, and the desired driving torque of the left and right driving wheels is output to the racing model built by CarSim. The simulation results are shown in Figures 15–19.

It can be seen from the simulation diagram of position tracking in Figure 15 and the simulation diagram of lateral position tracking error in Figure 16 that, except for the error of 0.2 m in sharp corners, the other position tracking errors are less than 0.1 m, as can be seen from the simulation diagram of steering wheel angle in Figure 17. It can be seen that the total rotation angle composed of the target rotation angle and the feedback rotation angle changes smoothly and meets the actual steering angle requirements. It can be seen from the simulation diagram of running speed in Figure 18 and the simulation diagram of left and right wheel drive torque in Figure 19 that the initial torque is positive. In an accelerating condition, the additional yaw moment is provided by the differential torque to compensate for the tracking error when understeer occurs at

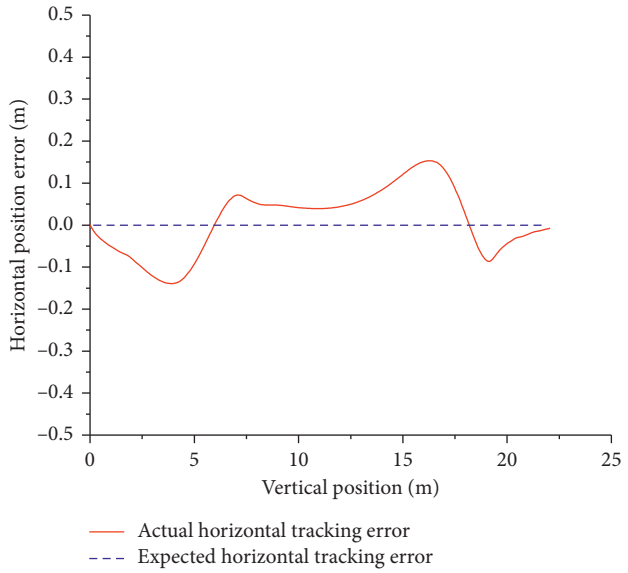


Figure 12: Simulation diagram of horizontal position tracking error.

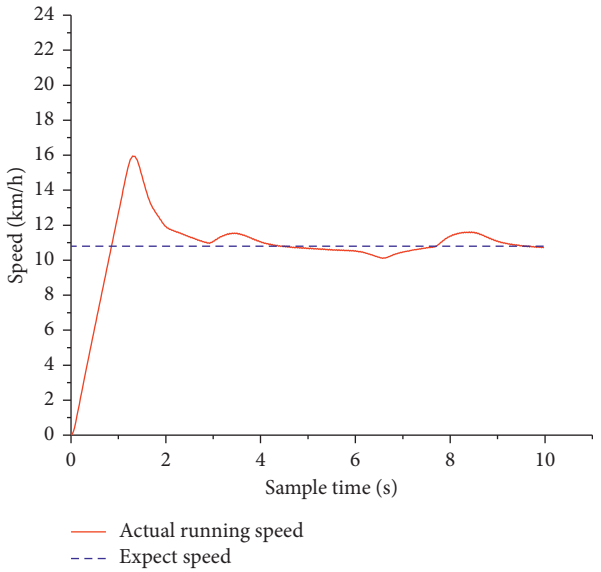


Figure 14: Curve tracking speed simulation diagram.

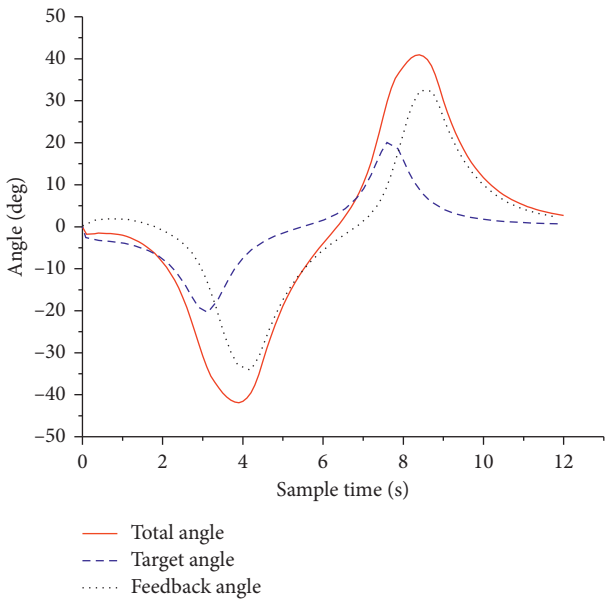


Figure 13: Simulation diagram of curve tracking steering angle.

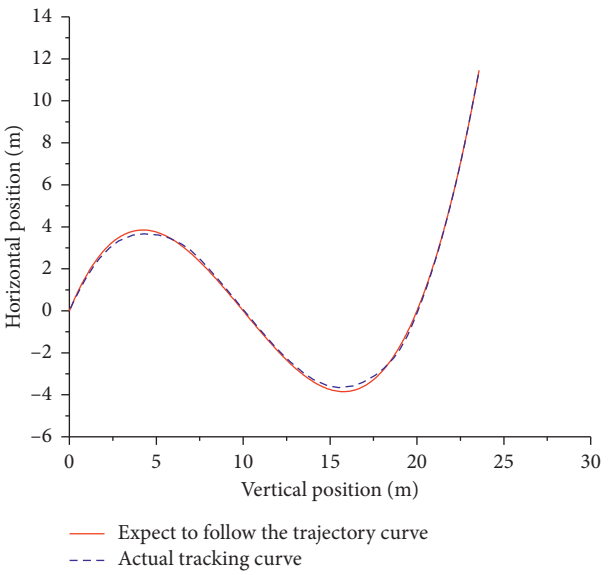


Figure 15: Position tracking simulation diagram.

the corner, and the driving torque on both sides is basically the same when it is close to a straight line. Therefore, the addition of the differential assisted steering control strategy trajectory tracking controller not only reduces the tracking error, but also assists the driverless racing car to pass the tight curve track smoothly.

5.4. Actual Car Test Verification. In order to verify the effect of the designed trajectory tracking controller, the actual car test verification analysis is carried out. Using Liaoning University of Technology's dual-motor-driven driverless racing car as an experimental platform, in the open and flat ground, a curve simulation working condition was designed according to the high-speed tracking project in the competition, and the trajectory tracking of the driverless racing car without differential drive was compared. In integrated control strategy for control and trajectory tracking of driverless cars, as shown in Figure 20, the driverless car carries on the actual car test according to the given track. According to the coordinate information of the cone and barrel located by the lidar, camera, and GPS, plan the desired tracking trajectory curve, which is shown by the visualization interface of the ROS system and the desired trajectory curve is shown in Figure 21.

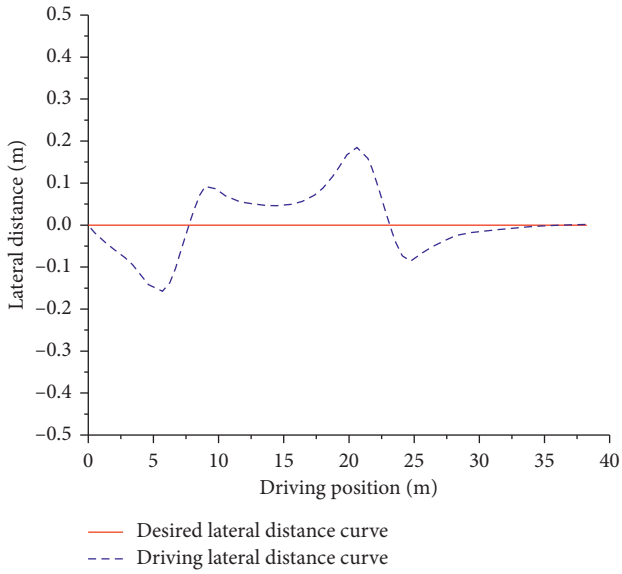


Figure 16: Simulation diagram of lateral position tracking error.

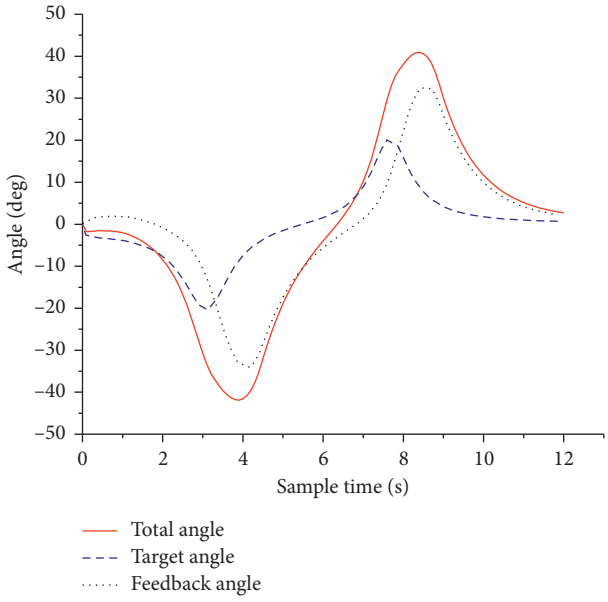


Figure 17: Simulation diagram of steering wheel angle.

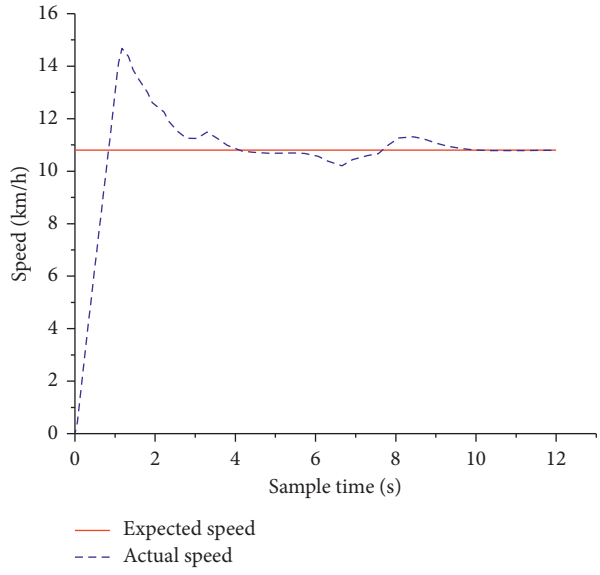


Figure 18: Running speed tracking simulation diagram.

The trajectory tracking control experiment of a driverless car without differential drive assisted steering is carried out. The comparison of actual car experimental data is shown in Figures 22–23.

The built trajectory tracking controller can ensure that the driverless racing car can track the expected trajectory well, except for the fluctuation of the expected trajectory curve curvature, which can be seen from Figures 22 and 23. In addition to the large deviation, the lateral error between the actual tracking trajectory and the expected tracking trajectory is about 15 cm, which fully meets the required tracking accuracy in the actual car verification. It can be seen from the experimental diagram of steering angle in Figure 24 that the target rotation angle and the total input angle formed by the feedback angle are relatively stable at other times except for large changes.

In order to track the desired trajectory as much as possible at large turns, which meets the actual conditions of

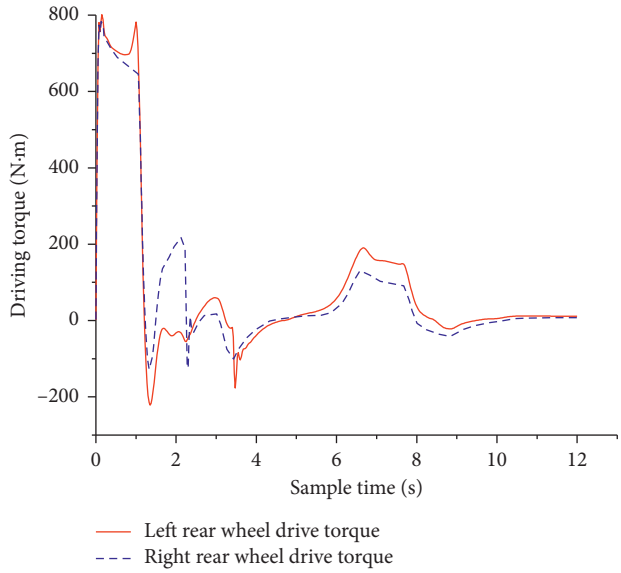


Figure 19: Left and right driving torque simulation diagram.

Figure 20: Actual car test diagram.

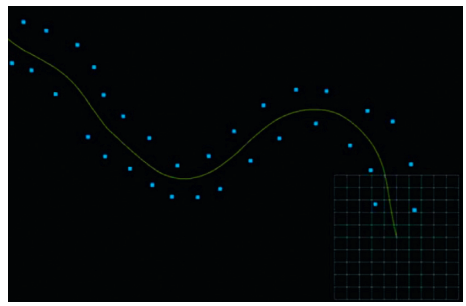


Figure 21: Expected trajectory curve.

the car, the feedback angle can always be corrected in time when there is a lateral tracking deviation. The target angle follows the desired trajectory curve.

The actual car test of the trajectory tracking control of the driverless racing car based on the differential assisted steering of the yaw rate is shown in Figures 25–29.

It can be concluded that the strategy which uses additional differential drive torque to assist steering control improves the accuracy of trajectory position tracking of driverless car, and the effect is more effective at big turns from Figures 25 and 26. Obviously, the average error is controlled at about 10 cm, which is less than the tracking error when the differential drive torque control strategy is not added, and the tracking accuracy is improved. It can be seen from Figure 27 that the speed fluctuates up and down the expected speed and finally can keep up with the expected speed. When the car turns to

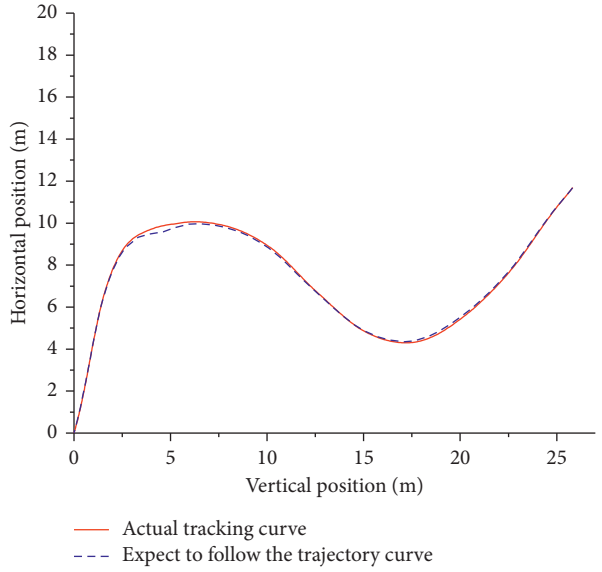


Figure 22: Trajectory position tracking comparison chart.

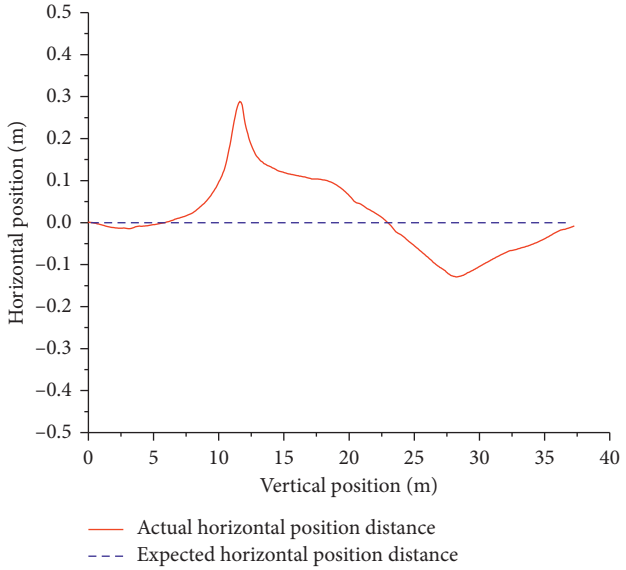


Figure 23: Comparison of horizontal tracking positions.

the right, the driving torque of the left rear wheel is greater than the driving torque of the rear wheel. The differential driving torque provides additional yaw moment for the car, which is used to reduce the steering wheel slip angle and understeer tendency and improve the car track accuracy.

It can be seen from Figure 28 that the total input angle of the desired trajectory changes greatly in order to track the desired trajectory as much as possible at a large turn, and the other time is relatively stable. It meets the actual conditions of the car, and when the lateral tracking deviation occurs, the feedback corner can always correct the target corner to track the desired trajectory curve in time. And when the lateral tracking deviation occurs, the feedback angle can always correct the target angle to track the desired trajectory curve in time.

It can be seen from Figure 29 that when the car turns to the right, the driving torque of the left rear wheel is greater than the driving torque of the rear wheel, and the resulting

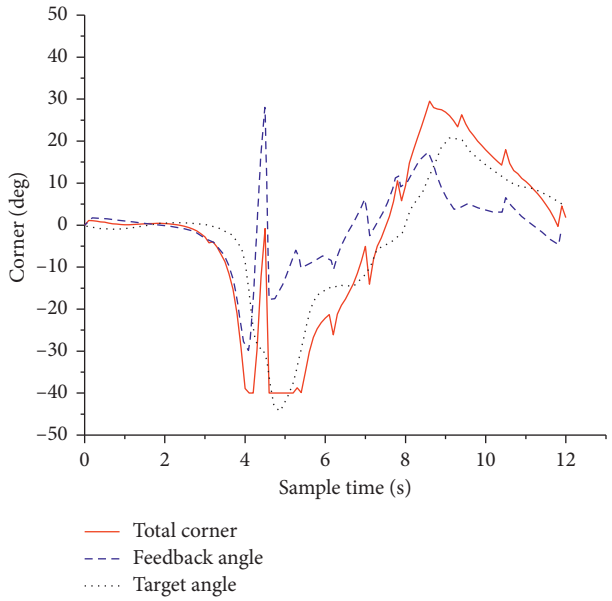


Figure 24: Experimental diagram of steering angle.

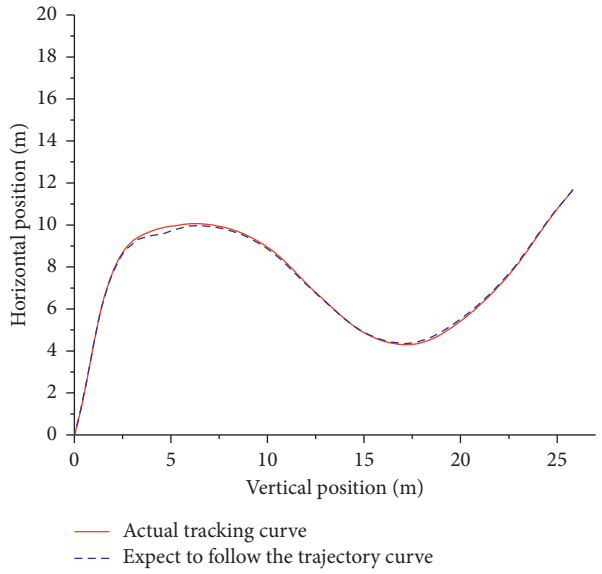


Figure 25: Trajectory position tracking.

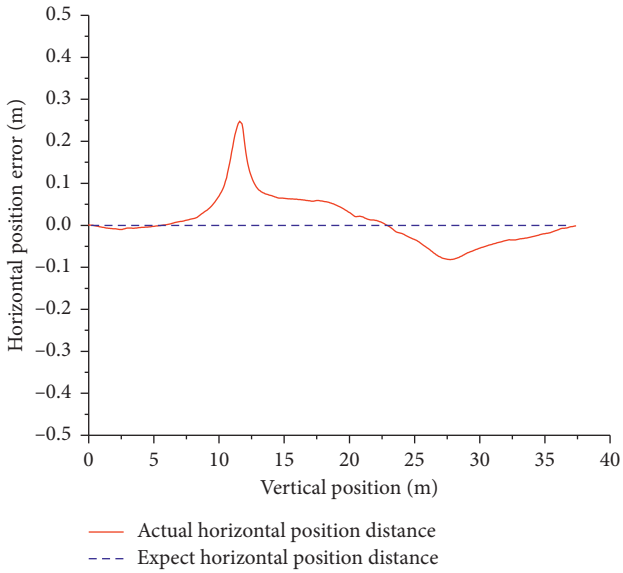


Figure 26: Comparison of horizontal tracking.

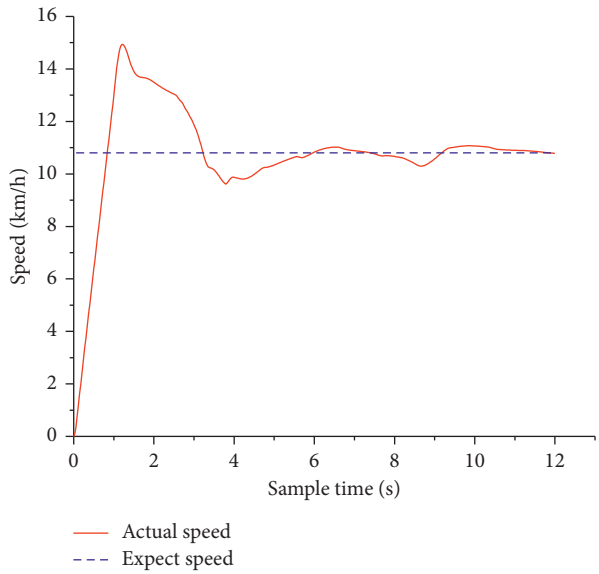


Figure 27: Comparison of horizontal tracking.

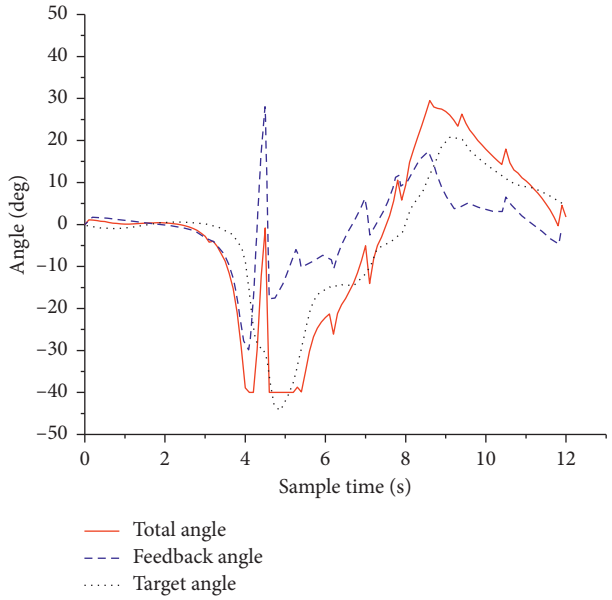


Figure 28: Steering angle analysis diagram.

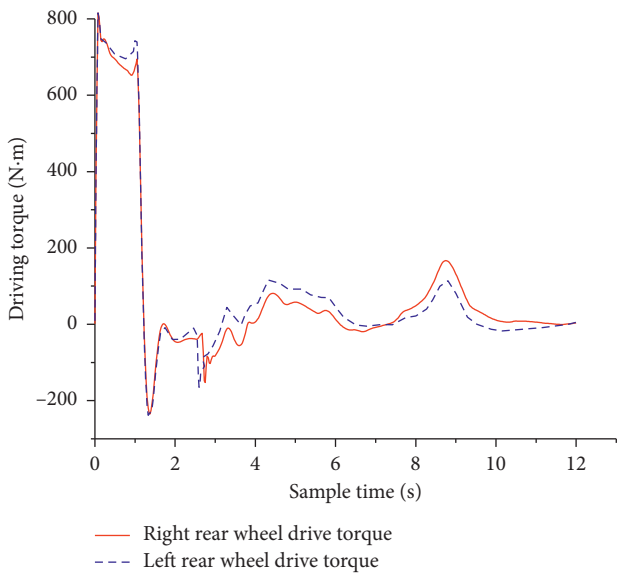


Figure 29: Left and right driving torque simulation diagram.

differential driving torque provides additional yaw moment for the car to reduce the steering wheel slip angle and insufficient steering trend improves the accuracy of racing track tracking.

6. Conclusion

(1) Aiming at improving the accuracy of trajectory tracking and the tracking timeliness, the tracking accuracy and tracking targets in the process of trajectory tracking of driverless racing car are analyzed. The optimal control theory is applied to design the LQR active steering controller of the driverless racing car. At the same time, the dual-motor-driven car is introduced to enable differential drive. Based on the fuzzy control theory, the differential drive control method is designed to improve the accuracy of estimating tracking degree.

(2) Simulation verification and actual car test verification results show that, under given conditions, the lateral tracking error of the trajectory tracking controller designed in this paper is 0.15 m and it can track quickly when the trajectory deviation occurs. After adding the differential drive control, the trajectory tracking control accuracy is improved by nearly 30% with strong accuracy and timeliness. The developed method can be applied in multiagent system [19] and complex pipeline networks [20] in the future.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Science Foundation of China (51675257 and 51305190) and Beijing Key Laboratory of Measurement & Control of Mechanical and Electrical System Technology, Beijing Information Science & Technology University (KF20202223201).

References

[1] R. Attia, R. Orjuela, and M. Basset, "Combined longitudinal and lateral control for automated vehicle guidance," *Vehicle System Dynamics*, vol. 52, no. 2, pp. 261–279, 2014.

[2] C. Fiho and D. Wolf, "Dynamic inversion-based control for front wheel drive autonomous ground vehicles near the limits of handling," in *Proceedings of the 2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, Qingdao, China, October 2014.

[3] N. R. Kapania and J. C. Gerdes, "Design of a feedback feedforward steering controller for accurate path tracking and stability at the limits of handling," *Vehicle System Dynamics*, vol. 53, no. 12, pp. 1–18, 2015.

[4] M. Brown, J. Funke, S. Erlien, and J. C. Gerdes, "Safe driving envelopes for path tracking in autonomous vehicles," *Control Engineering Practice*, vol. 61, no. 3, pp. 307–316, 2017.

[5] J. Shah, M. Best, A. Benmimoun, and M. L. Ayat, "Autonomous rear-end collision avoidance using an electric power steering system," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 229, no. 12, pp. 1638–1655, 2015.

[6] S. Inoue, T. Ozawa, H. Inoue et al., "Cooperative lateral control between driver and ADAS by haptic shared control using steering torque assistance combined with Direct Yaw Moment Control," in *Proceedings of the 2016 IEEE International Conference on Intelligent Transportation Systems*, Rio de Janeiro, Brazil, November 2016.

[7] U. Rosolia, S. D. Bruyne, and A. G. Alleyne, "Autonomous vehicle control: a nonconvex approach for obstacle avoidance," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 2, pp. 1–16, 2016.

[8] J. Su, J. Wu, P. Cheng et al., "Autonomous vehicle control through the dynamics and controller learning," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 9712–9726, 2018.

[9] X. Ji, J. Wang, Y. Zhao, Y. Liu, L. Zang, and B. Li, "Path planning and tracking for vehicle Parallel parking based on preview BP neural network PID controller," *Transactions of Tianjin University*, vol. 21, no. 3, pp. 199–208, 2015.

[10] X. Ji, X. He, C. Lv, Y. Liu, and J. Wu, "Adaptive-neural-network-based robust lateral motion control for autonomous vehicle at driving limits," *Control Engineering Practice*, vol. 76, no. 3, pp. 41–53, 2018.

[11] H. Guo, J. Liu, D. Cao et al., "Dual-envelop-oriented moving horizon path tracking control for fully automated vehicles," *Mechatronics*, vol. 50, no. 3, pp. 41–53, 2018.

[12] J. Ni and J. H. C. Hu, "Envelope control for four-wheel independently actuated autonomous ground vehicle through AFS/DYC integrated control," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 9712–9726, 2017.

[13] K. Liu, J. Gong, S. Chen, Y. Zhang, and H. Chen, "Model predictive stabilization control of high-speed autonomous ground vehicles considering the effect of road topography," *Applied Sciences*, vol. 8, no. 5, pp. 822–837, 2018.

[14] W. Chen, J. Wang, M. Wang, and J. Wang, "Adaptive preview control of lateral movement of intelligent vehicle with visual navigation," *China Mechanical Engineering*, vol. 25, no. 5, pp. 698–704, 2014.

[15] G. Lie, Z. Yi-bing, L. Lin-hui, G. Ping-shu, and H. Xiao-hui, "Antisideslip and antirollover safety speed controller design for vehicle on curved road," *Mathematical Problems in Engineering*, vol. 2014, no. 3, 12 pages, Article ID 253176, 2014.

[16] L. Liu, Y.-J. Liu, A. Chen, S. Tong, and C. L. Philip Chen, "Integral barrier Lyapunov function based adaptive control for switched nonlinear systems," *Science China Information Sciences*, vol. 63, no. 3, Article ID 132203, 2020.

[17] R. Wang, Q. Sun, W. Hu, Y. Li, D. Ma, and P. Wang, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, pp. 7866–7879, 2021.

[18] X. Hu, H. Zhang, D. Ma, and R. Wang, "A tnGAN-based leak detection method for pipeline network considering incomplete sensor data," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.

[19] Q. Sun, R. Han, H. Zhang, J. Zhou, and J. M. Guerrero, "A multiagent-based consensus algorithm for distributed coordinated control of distributed generators in the energy internet," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 3006–3019, 2015.

[20] D. Ma, Y. Li, X. Hu, H. Zhang, and X. Xie, "An optimal three-dimensional drone layout method for maximum signal coverage and minimum interference in complex pipeline networks," *IEEE Transactions on Cybernetics*, 2021, In Press.

Hindawi
Complexity
Volume 2021, Article ID 9995033, 12 pages
https://doi.org/10.1155/2021/9995033

Research Article

A Pinning Actor-Critic Structure-Based Algorithm for Sizing Complex-Shaped Depth Profiles in MFL Inspection with High Degree of Freedom

Zhenning Wu,[1,2] **Yiming Deng**,[2] **and Lixing Wang**[3]

[1]*College of Information Science and Engineering, Northeastern University, Shenyang 110004, China*
[2]*Department of Electrical and Computational Engineering, Michigan State University, East Lansing, MI 48823, USA*
[3]*College of Computer Science and Engineering, Northeastern University, Shenyang 110004, China*

Correspondence should be addressed to Zhenning Wu; wuzn2003@hotmail.com

Received 14 March 2021; Revised 1 April 2021; Accepted 15 April 2021; Published 24 April 2021

Academic Editor: Rui Wang

One of the most efficient nondestructive methods for pipeline in-line inspection is magnetic flux leakage (MFL) inspection. Estimating the size of the defect from MFL signal is one of the key problems of MFL inspection. As the inspection signal is usually contaminated by noise, sizing the defect is an ill-posed inverse problem, especially when sizing the depth as a complex shape. An actor-critic structure-based algorithm is proposed in this paper for sizing complex depth profiles. By learning with more information from the depth profile without knowing the corresponding MFL signal, the algorithm proposed saves computational costs and is robust. A pinning strategy is embedded in the reconstruction process, which highly reduces the dimension of action space. The pinning actor-critic structure (PACS) helps to make the reward for critic network more efficient when reconstructing the depth profiles with high degrees of freedom. A nonlinear FEM model is used to test the effectiveness of algorithm proposed under 20 dB noise. The results show that the algorithm reconstructs the depth profile of defects with good accuracy and is robust against noise.

1. Introduction

Magnetic flux leakage (MFL) is one of the most widely used NDT techniques, which has been widely used for inspection of oil and gas pipeline since the 1960s. It is efficient in finding defects caused by corrosion and mechanical damage and other metal loss defects for pipelines and storage tanks [1–3]. It is helpful to give the health condition of the working facilities to the operators, which prevents disasters to environment, industry, and human being due to the leakage of explosive or dangerous chemicals. Estimating the shape of defect is the key problem of inspection. Though MFL is efficient in finding the defects and anomalies, the reconstruction process from inspection signal to defect depth is not an easy task, as it is usually contaminated by sampling noise [4]. Reconstruction results with more details such as the detailed shape of the defect rather than only length, width, and depth are more helpful to estimate the health condition of the tested material [5]. Among length, width, and depth, depth reconstruction is the most challenging part as it is highly ill-posed. Unfortunately, reconstructing the defect shape with details makes the ill-posed inverse problem even harder to be solved.

The solutions of MFL inverse problem could be classified as either non-model-based methods or model-based methods. Non-model-based methods solve this inverse problem by building a mapping between sampled signal and the shape of defect. Neural networks are usually used to build this mapping [6–10]. The input of this neural network can be the signal of MFL inspection, and the sizing information of the defect is set as the output. These methods are fast but highly rely on the data set used to train the neural network. The accuracy is highly impacted by the quality of training data set.

A forward physical model is involved in the model-based methods. The forward model is used to give simulated signal to a given depth profile. The simulated signal is used for comparison with the reference signal. The residual error between simulated signal and reference signal is used to give information for the iteration strategy. By minimizing the residual error, the size of the defect is repetitively computed [11–16]. Numerical models and analytical models are two categories of methods involved as forward model of MFL. Analytical models are fast but have more limitations as the model is derived with many simplifications, making it less accurate [17, 18]. Numerical model provides accurate results, but it is computationally very costly especially when a fine model is needed. The design of the iteration policy for numerical model is another problem that is hard to design. Classic methods design the policy with gradient information to minimize the residual error [19, 20]. These methods usually have some limitations to assume the shape of the defect a priori. Another kind of solution uses a mapping which is trained to replace the numerical forward model. A novel iterative method of inversion using adaptive wavelets and radial basis function neural network are proposed in [5]. A RBF neural network is used as a forward model in [21]. Heuristic methods are the third kind of solution of designing the iteration policy. Han et al. proposed a particle swarm optimization method to solve this problem [22]. Li et al. proposed a modified harmony search algorithm as the iteration policy [23]. As these heuristic methods are not deterministic, they usually need a vast amount of forward model evaluations.

Considering the state-of-the-art solutions, there are still some common problems in solving the problem of sizing the defects. First, for the non-model-based method, the mapping is trained according to the data without exploration to data not included in the training set. It makes the mapping highly rely on the distribution of the training data set. As the MFL inverse problem is ill-posed, the mapping from signal to defect profiles can also be troubled by the nonuniqueness of the mapping. Second, for the model-based method, the iteration strategy is designed based on the forward model in use and highly relies on it. For numerical model, it has high performance in simulating the inspection signal, but it is hard to build an iteration strategy based on it.

The similarity between the state-of-the-art machine learning technique of game play [24–29] and classic model-based iteration method inspires the study in this paper. Though the RL algorithm is basically a machine learning technique which needs training, the general structure has similarity with the classic iteration method which makes it possible to design an iteration strategy for numerical forward physics model. An actor-critic structure is adopted to design the iteration strategy. The actor network gives the iteration strategy, and the performance is evaluated by the critic network which improves the strategy given by actor network in the coming steps. For the problem with high dimensions in its action space, the "reward" which is used to improve the performance by critic network does not perform as efficient as it does with the problem of lower action space dimensions [30–32]. A pinning-based strategy is given in this paper to reduce the dimension of action space, which helps to make the critic network more efficient.

The principle of actor-critic based structure is introduced in Section 2 along with the principle of MFL inspection. The detail of PACS algorithm is described in Section 3. Simulated inspection signal from a nonlinear FEM model is used to test the performance of the algorithm proposed under 20 dB noise in Section 4. The conclusion is drawn in Section 5.

2. Model and Principle

2.1. Physics Model. The principle of MFL inspection is based on electromagnetic theory. By magnetizing the test material into saturation, a magnetic flux leakage can be detected by Hall-effect sensors where a defect is located. Strong permanent magnets are usually used to magnetize the testing material. The Hall-effect sensors are usually located close to the surface of the tested material. The magnetizing and sensing principles are illustrated in Figure 1.

The principle of MFL inspection is magnetic, where Maxwell's equation can be used to describe its behavior:

$$\nabla \times \left(\mu^{-1} \nabla \times \mathbf{A}\right) = \mathbf{J}, \tag{1}$$

where μ, $\mathbf{J}$, and $\mathbf{A}$ represent the permeability of the media, the source current density, and the magnetic vector potential. In (1), μ is usually not a constant due to the property of the material and can be described as a function of magnetic flux density $\mathbf{B}$ as $\mu = \mu(\mathbf{B})$. The magnetic flux density which can be collected with the Hall-effect sensors is $\mathbf{B} = \nabla \times \mathbf{A}$. For a simple defect, the magnetic flux density $\mathbf{B}$ is illustrated in Figure 2. The x-axis component is usually sampled as the inspection signal. There are two ways to solve this Maxwell's equation formulated as (1). Numerical solution such as finite element method (FEM) is widely used to solve this partial differential equation. Another kind of solution is dipole model, which makes some simplification of the forward model, which gives an analytical solution to this forward model [19, 33, 34].

As the analytical solution cannot provide enough accuracy for complex defects, numerical methods are usually used to get a numerical solution for these problems. FEM is a widely used method to get a numerical solution for partial differential equation. The general process of a FEM solution is as follows. First, the partial differential equation is transformed into corresponding variational functional equation. Then, the domain that needs to be computed is divided into certain number of finite elements. By assembling all the variational functional equations of all the elements within the domain, the solution can be obtained by solving

$$\mathbf{K} \cdot \mathbf{a} = \mathbf{f}, \tag{2}$$

where $\mathbf{a}$ represents the nodal solutions of elements of discrete approximation in the form of a vector, $\mathbf{K}$ is the sparse element stiffness matrix, and $\mathbf{f}$ is the source vector containing the boundary conditions and model inputs.

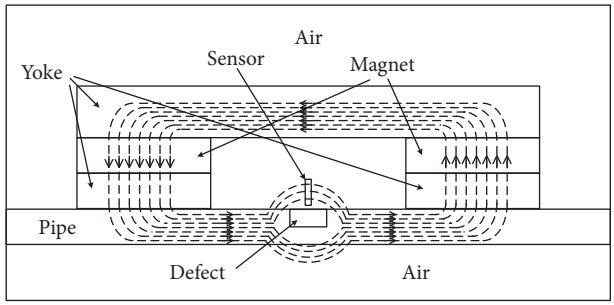


Figure 1: Principle of MFL inspection.

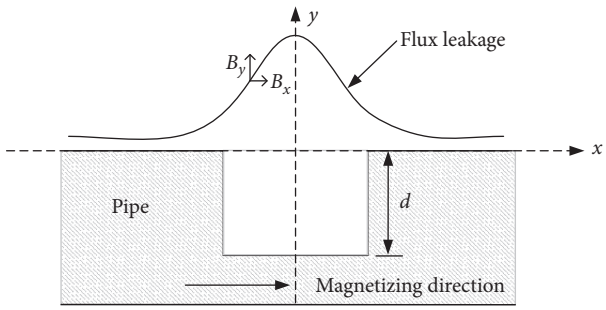


Figure 2: Illustration of magnetic flux leakage of a simple defect.

Commonly, the MFL inspection model is built with components as described in Figure 1. As the senor position is fixed with the magnetizing components, it needs to rebuild the defect location n times if a sequence of data with n points are sampled. It means to repeat the forward model n times, which is computationally very costly. A simplified model is adopted in this paper, which is proposed in [14]. The simplified model is illustrated in Figure 3. As the principle of MFL inspection is magnetizing the material into saturation, a pair of paralleled current layers are adopted to magnetize the testing material. The commonly used permanent magnets and yokes are removed. By adopting this model, the principle and nonlinear character are kept. A sampled signal with n sampling points only needs to run the forward model once, which saves a lot computationally. The region of interest (ROI) shown in Figure 3 represents the domain where the defect is going to be reconstructed by the algorithm proposed. The depth profile within ROI consists of several subdefects, which make up a complex depth profile.

2.2. Principle of RL. Reinforcement learning considers the paradigm of an agent interacting with its environment aiming to learn a behavior which maximizes the reward. The agent consists of an actor network and a critic network. The actor network is trained to decide which action should be taken at current state. The critic network evaluates each action based on its current state with reward and improves the strategy of the actor network. There are four definitions in RL, state x_t, reward r_t, action a_t, and environment E. An agent takes action a_t at current state x_t, where t is discrete time-step. The action a_t interacts with certain environment

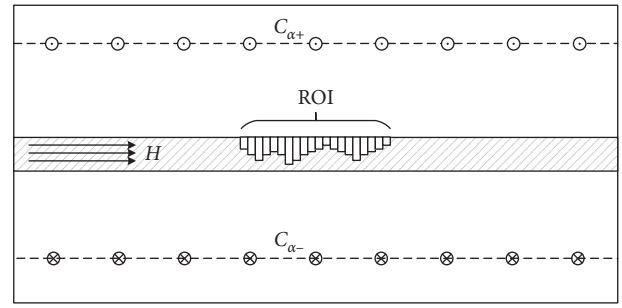


Figure 3: Illustration of simplified FEM model used in this paper with complex depth profile.

and obtains a new state x_{t+1} with a reward r_t, which evaluates the performance of this action. The action is defined by a deterministic policy π: $\mathcal{S} \longrightarrow \mathcal{A}(\mathcal{A})$ which is a mapping from states to actions. $\mathcal{S}$ represents the state space, and $\mathcal{A}$ represents the action space. A discounted sum of future rewards is called a return as $R_t = \sum_{i=t}^{T} \gamma^{(i-t)} r(\mathbf{s_i}, \mathbf{a_i})$, where $\gamma \in [0, 1]$ is a discounting factor. The agent's goal is to maximize the expected returns $\mathbb{E}_\pi[R_1]$. The action-value function is defined as $Q^\pi(\mathbf{s_t}, \mathbf{a_t}) = \mathbb{E}_\pi[R_t|\mathbf{s_t}, \mathbf{a_t}]$. It is also called Q-function, which represents the expected return after taking an action $\mathbf{a_t}$ in state $\mathbf{s_t}$ and thereafter following the policy π. The critic is updated by minimizing the loss L as follows:

$$L = \frac{1}{N} \sum_i \left(y_i - Q\left(s_i, a_i|\theta^Q\right)\right)^2, \tag{3}$$

where

$$y_i = r_i + \gamma Q'\left(s_{i+1}, \pi'\left(s_{i+1}|\theta_{\pi'}\right)|\theta^{Q'}\right). \tag{4}$$

The actor policy is updated using sampled gradient as follows:

$$\nabla_{\theta^\pi}\pi|_{s_i} \approx \frac{1}{N} \sum_i \nabla_a Q\left(s, a|\theta^Q\right)|_{s=s_i, a=\pi(s_i)} \nabla_{\theta_\pi}\pi\left(s|\theta^\pi\right)|_{s_i}. \tag{5}$$

In the problem of sizing the depth profile of MFL inspection, four parameters are involved in the reconstruction process, the reference depth profile $\mathbf{d}_{\text{ref}}$, reconstructed depth profile $\mathbf{d_t}$ at time-step t, reference signal $\mathbf{s}_{\text{ref}}$, and signal of reconstructed depth profile $\mathbf{s_t}$ at time-step t. Reference depth profile is the target of the reconstruction, which is not observable during the entire reconstruction process. The other three parameters are fully observable all the time. During training of non-model-based method, the reference depth profile and corresponding reference signal are used to train the mapping. The model-based method only involves the reference signal during the reconstruction process. In this paper, three parameters, reference depth profile, reference signal, and reconstructed depth profile, are used to train the actor-critic structure-based algorithm proposed. The involvement of reconstructed depth profile gives more information of the depth profile space. The signal of reconstructed signal is not utilized as it is computationally costly.

This paper is inspired by the similarity between the training process of RL and model-based iteration method. The similarity is illustrated as shown in Figure 4. For the iteration structure, it starts with an initial defect depth profile. According to the iteration strategy, a reconstructed depth profile is given. The signal of corresponding depth profile is generated with the forward model. By comparing the signal of reconstructed depth profile with reference signal, a residual is obtained. This process iterates until the residual is smaller than a threshold when a final reconstructed depth profile is obtained. The learning process of the actor-critic structure-based RL method proposed in this paper has similarities with the model-based solution mentioned above. The state can be that the depth profile needs to be reconstructed in a certain form. The agent is the strategy controlling the iteration process. Action is the output of the agent, which controls how to change the state until termination criteria are satisfied. The strategy that controls the iteration process is learned from the data given and generated during the iteration process. It solves the problem that it is hard to design iteration strategy for numerical forward models. The performance of the strategy is evaluated by the critic network with rewards from each step. By involving the data generated during the iteration process, more data is given, which means the training data is not limited to the given data pairs of depth profile and corresponding signal. It improves the problem that the non-model-based solution highly relies on the distribution of the training data.

3. Algorithm

In this paper, an actor-critic structure-based RL method for complex depth profile reconstruction is proposed. The algorithm of Deep Deterministic Policy Gradients (DDPG) is adopted to train the actor-critic structure [35–37]. The definitions of parameters for the problem of sizing the depth profile of MFL inspection are described as follows. The state is defined as $\mathbf{s_t} = (\mathbf{d_t}, \mathbf{s}_{\text{ref}})$, which consists of two parts, the normalized reconstructed depth profile and the reference signal. $\mathbf{d_t}$ is the normalized reconstructed depth at time-step t. $\mathbf{s}_{\text{ref}}$ is the sampled reference signal. As the signal outside the ROI has less characteristics than signal within the domain of ROI, most of the sampling points are selectively removed to reduce the state dimension. The change of $\mathbf{d_t}$ at each time-step is taken as action. Different from the model-based method that uses the residual between reference signal and signal of reconstructed depth profile to evaluate the performance, the performance of the actor network is evaluated with reward at each time-step. Reward is designed as minus value of Euclidean distance between reference depth profile and reconstructed depth profile as shown in the following equation:

$$r_t = -\left\|\mathbf{d}_{\mathbf{t}+1} - \mathbf{d}_{\text{ref}}\right\|. \tag{6}$$

As the target of this MFL inverse problem is sizing the depth profile as precisely as possible, it means the subdefect $d(i)$ needs to approach its corresponding reference subdefect with small error. Then, the complexity is associated

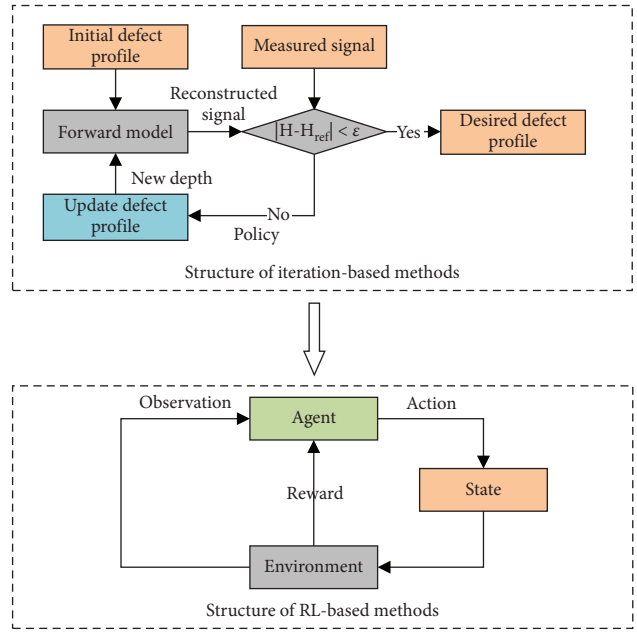


Figure 4: Similarity between the learning process of reinforcement learning and classic iteration method.

with the dimension of the degree of freedom of this inverse problem. The problem when encountering a high-dimensional degree of freedom comes that the reward that evaluates the performance of the actor network becomes less efficient. It is because the measurement of distance becomes less efficient in high-dimensional problems [30–32]. In this paper, instead of giving each subdefect an action to control its depth, limited subdefects are selected to accept action given by the actor network. Subdefects accepting control from actor network are called pinning subdefects in this paper. As the defects are usually caused by corrosion or mechanical damage, the difference between adjoining subdefects is usually not sharp. It means that it is possible to use some pinning subdefects, which are controlled by the action to represent the depth profile. Subdefects between two pinning subdefects are interpolated. By adopting this pinning strategy, the dimension of action space is reduced significantly. The reward is still calculated using the full information of depth profile. As the dimension of action space is reduced, the measurement of distance becomes more efficient than using full space. The entire depth profile with pinning subdefects is illustrated in Figure 5. The flowchart of PACS learning process within one episode is illustrated in Figure 6. The entire algorithm including the pinning strategy with the learning process of the actor-critic structure-based reconstruction algorithm is given in Algorithm 1.

From Algorithm 1, it can be seen that, within one episode, $\mathbf{s}_{\text{ref}}$ as part of the states is combined with many depth profiles $\mathbf{d_t}$ generated during the iteration process. It means that, despite the relationship between ultimate reconstructed depth profile and reference signal, the function in the depth profile space is also explored by PACS proposed in this paper. It helps to have better

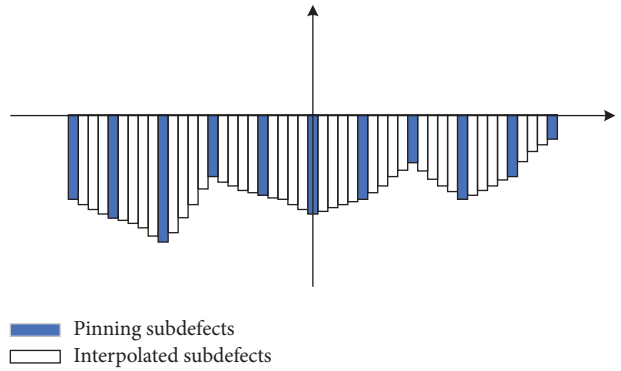


Figure 5: Illustration of pinning-based depth reconstruction within the ROI.

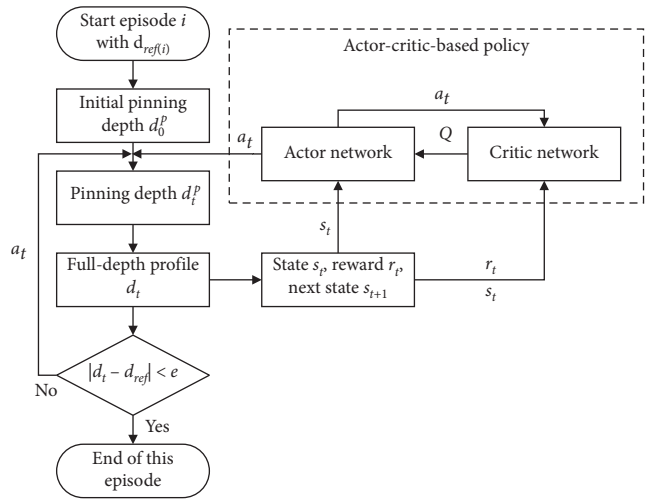


Figure 6: Flowchart of PACS learning process within one episode.

(i) Initialize actor network $\mu(\mathbf{s}|\theta^{\pi})$, critic network $Q(\mathbf{s}, \mathbf{a}|\theta^{Q})$, target network π' and Q', replay buffer R
(ii) For episode = 1, M do
(iii) Initialize pinning subdefects $\mathbf{d_0^p}$, interpolate to have the full depth profile
(iv) Get initial observation state $\mathbf{s}_1$ from reference signal and depth of sub-defects
(v) For $t = 1$, T do
(vi) Generate an action $\mathbf{a_t} = \mu(\mathbf{s_t}|\theta^{\pi}) + \mathcal{N}_\mathbf{t}$ from the output of actor network and exploration noise process
(vii) Execute action $\mathbf{a_t}$, obtain new depth of pinning sub-defects $\mathbf{d_t^p}$
(viii) Interpolate to get the full depth profile $\mathbf{d_t}$ within the ROI, calculate reward $\mathbf{r_t}$ and new state $\mathbf{s_{t+1}}$
(ix) Store $(\mathbf{s_t}, \mathbf{a_t}, r_t, \mathbf{s_{t+1}})$ in R
(x) If capacity of replay buffer R is full then
(xi) Randomly sample N piece of data $(\mathbf{s_t}, \mathbf{a_t}, r_t, \mathbf{s_{t+1}})$ from R
(xii) Update the critic network and actor network with (5) and (3)
(xiii) Update the target networks:
(xiv) $\theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta_{Q'}$
$\theta^{\pi'} \leftarrow \tau\theta^{\pi} + (1-\tau)\theta_{\pi'}$
(xv) end if
(xvi) If error between each reference subdefect and reconstructed subdefect is less than ε, then
(xvii) break
(xviii) end if
(xix) end for
(xx) end for

Algorithm 1: PACS.

reconstruction results in an optimal way. As only component $\mathbf{d_t}$ in state $\mathbf{s_t} = (\mathbf{d_t}, \mathbf{s}_{\text{ref}})$ is updated, no forward model is called during one episode. It saves a lot as the call of forward model is computationally very costly if a fine model is required.

4. Results and Discussion

4.1. Model and Error Definitions. To test the accuracy of the algorithm proposed along with robustness, a simplified nonlinear numerical forward model is adopted as in [14]. The detail of the forward model is illustrated in Figure 3 and the dimension of the model can be found in Figure 7. There are 49 subdefects within the ROI. The adjoining subdefects stay tight and the span between centers of adjoining two subdefects is 2 mm. 11 subdefects are selected as pinning subdefects. The position of each pinning subdefect is shown in Figure 5. Subdefects between two pinning subdefects are interpolated with cubic interpolation. The x-component of the signal sampled with lift-off value of 1 mm above the surface is adopted as reference signal. The current density carried in the paralleled layer is $40000kA/m^2$ with opposite direction. The material is set as 1010 cold rolled steel. The property of the material including the B–H curve can be found in [14].

In order to test the effectiveness of the algorithm proposed in this paper, three error measurements are given. These measurements are root mean squared error (RMSE), peak depth error (PDE), and maximum deviation (MD). These measurements are described in (7)–(9) and illustrated in Figure 8. $d_{\text{ref}}(i)$ and $d(i)$ are the ith depths of subdefect for reference depth profile and reconstructed depth profile, respectively. From (7)–(9) and illustration in Figure 8, the error definitions can be understood in an easy way. MSE is commonly used in error measurements. PDE is the error between the maximum depths of reconstructed depth profile and reference depth profile. The subdefects from the reconstructed subdefect and the reference subdefect may not come from the same location within ROI. The value of 0.1 with regard to PDE means 1 mm error between peak depths if the wall thickness is 10 mm. MD is the maximum error between reconstructed subdefect and reference subdefect with the same location of the subdefects. The subdefects used to calculate MD value may not be the maximum depth of neither reconstructed depth profile nor reference depth profile. The value of 0.1 with regard to MD means 1 mm error if the wall thickness is 10 mm.

$$\text{RMSE}(d) = \sqrt{\frac{1}{N}\sum_{\text{ROI}}\left(d(i) - d_{\text{ref}}(i)\right)^2}, \tag{7}$$

$$\text{PDE}(\mathbf{d}) = \frac{1}{D}\left(\max\left(\mathbf{d}_{\text{ref}}\right) - \max\left(\mathbf{d}\right)\right), \tag{8}$$

$$\text{MD}(\mathbf{d}) = \frac{1}{D}\max\left(d_{\text{ref}}(i) - d(i)\right). \tag{9}$$

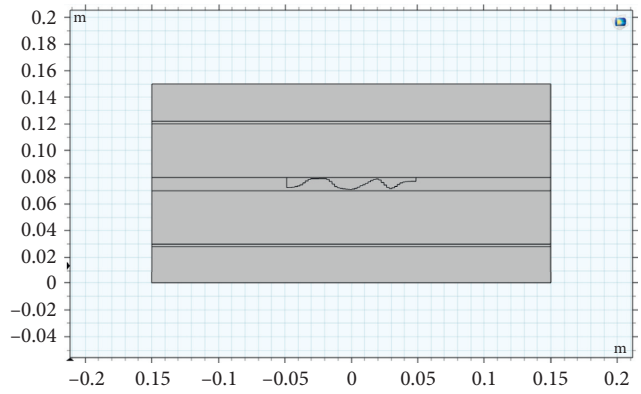


Figure 7: Dimension of the forward model with a defect of 49 subdefects.

4.2. Computing Results. The structure of actor network and critic network of PACS algorithm is described as follows. There are 82 neutrons in the input layer of actor network: 11 neutrons for the pinning subdefect normalized depth vector and 71 neutrons for the normalized reference signal. The sampling position of signal used as part of the state is illustrated in Figure 9. In the first hidden layer, there are 128 neutrons, and in the second hidden layer, there are 80 neutrons. 11 neutrons are set as the output layer, which controls the action for pinning subdefects. Besides, the activation function of output layer is set as "tanh", and all the activation functions are set as "ReLU." The input layer of critic network has two separated parts: one is 82 neutrons as the input layer of actor network and the other is 11 neutrons for the action. There are 128 neutrons, respectively, in the first hidden layer of critic network connecting the corresponding parts. The number of neutrons for the second and third hidden layer is 50, respectively. The output of critic network is Q value with one neutron.

The number of episodes M is 5000 with 200 time-steps for each episode as T. The size of replay buffer R is 1000000. The number of pieces of of sampling data at each time-step N is 128. The stop criterion ε is 0.5. The soft updating parameter τ is 0.01. The discounting factor γ is 0.1.

There are 10000 randomly generated complex defects with corresponding sampled signal generated with COMSOL Multiphysics 5.3a with MATLAB. 5000 of the pieces of data are used as training data set and 5000 others as testing data set. The algorithm proposed is coded with Python and TensorFlow 1.15. All the data and algorithm are run on a laptop with Intel i7 10750H processor and 16 GB RAM.

From error definition (7)–(9), the reconstructed results can be shown in different aspects. The results are shown with selected reconstruction results from the different ranking with MD value. The results of the 10%, 30%, 70%, and 90% are shown in Figure 10. The MD value is sorted from the smallest to the largest, which means Figure 10(a)is the best result among all the tested results shown in Figure 10 from the aspect of MD value. The ranking of the results in Figure 10 is tested using signal with SNR = 20 dB. Corresponding reconstructed results with noise-free signal for each depth profile and reference depth profile are also

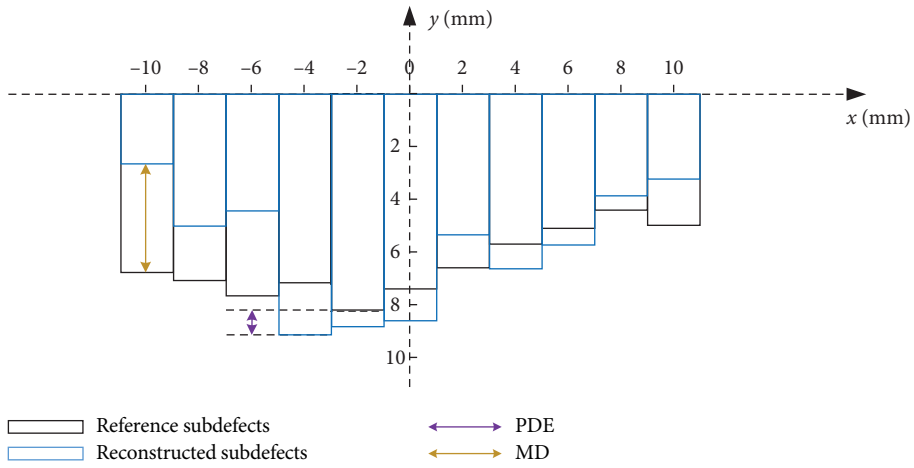


Figure 8: Illustration of PDE and MD.

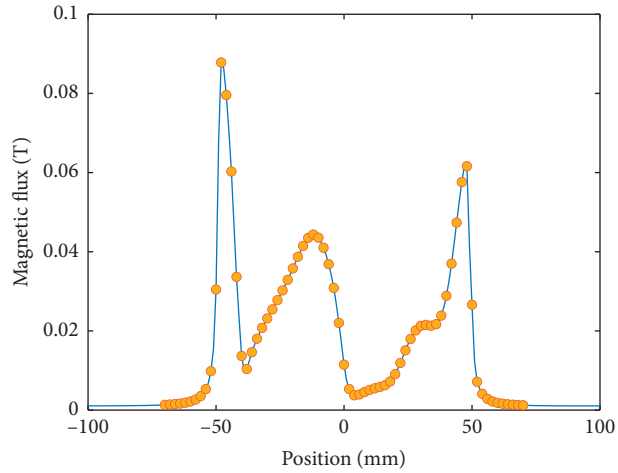


Figure 9: Illustration of sampling position of the inspection signal as part of state.

plotted in Figure 10. The corresponding values with ranking of each error definition are listed at the bottom of each subfigure. From Figure 10, it can be seen that all the reconstructed results follow the depth profile well. The worst result from PDE is in Figure 10(a) with the PDE ranking of 80.7% and value of -0.0786. The worst result from RMSE is in Figure 10(d) with the RMSE ranking of 87.2% and value of 0.1461. These are relatively small errors, proving the accuracy of the algorithm proposed in this paper. From the results in Figure 10, it can also be seen that the reconstructed results from signal of 20 dB noise are close to the results reconstructed from noise-free signal, which means the algorithm proposed in this paper is robust against noise.

The signals for corresponding reconstructed depth profile in Figure 10 are shown in Figure 11. The noise-free signal and signal with 20 dB noise are plotted in different color. The values of three error measurements are plotted in Figure 12 for the first 20 times-steps of reconstruction. They are results generated from reconstructing process of results in Figure 10 with noise signal of 20 dB. It can be seen from Figure 12 that all the error measurements converge in less than 10 time-steps. The results in Figure 12 show that the algorithm proposed in this paper converges fast with limited steps to the final reconstruction results.

To show the robustness of the algorithm proposed in this paper, the algorithm is trained with different size of training data sets. The size of the testing data set is 5000, which is the same as that in Figure 10. The sizes of training data sets are set as 2000, 3000, 4000, and 5000, respectively. A 20 dB noise is also added to the testing data. The results are shown as error distribution in Figure 13. The error distributions of results are shown as histogram. To make the figure clear, they are plotted with markers in different color instead of bars. The y-axis value of the markers represents the probability that the results fall into corresponding span. The width of the span is equal to that of the span between two adjoining markers from x-axis direction. The marker is located at the center of its corresponding span. The result is better if the error distributions are more concentrated and closer to zero. From Figure 13, it can be seen that, in contrast to the results from training with 2000 pieces of data, results from 3000, 4000, and 5000 have similar error distributions.

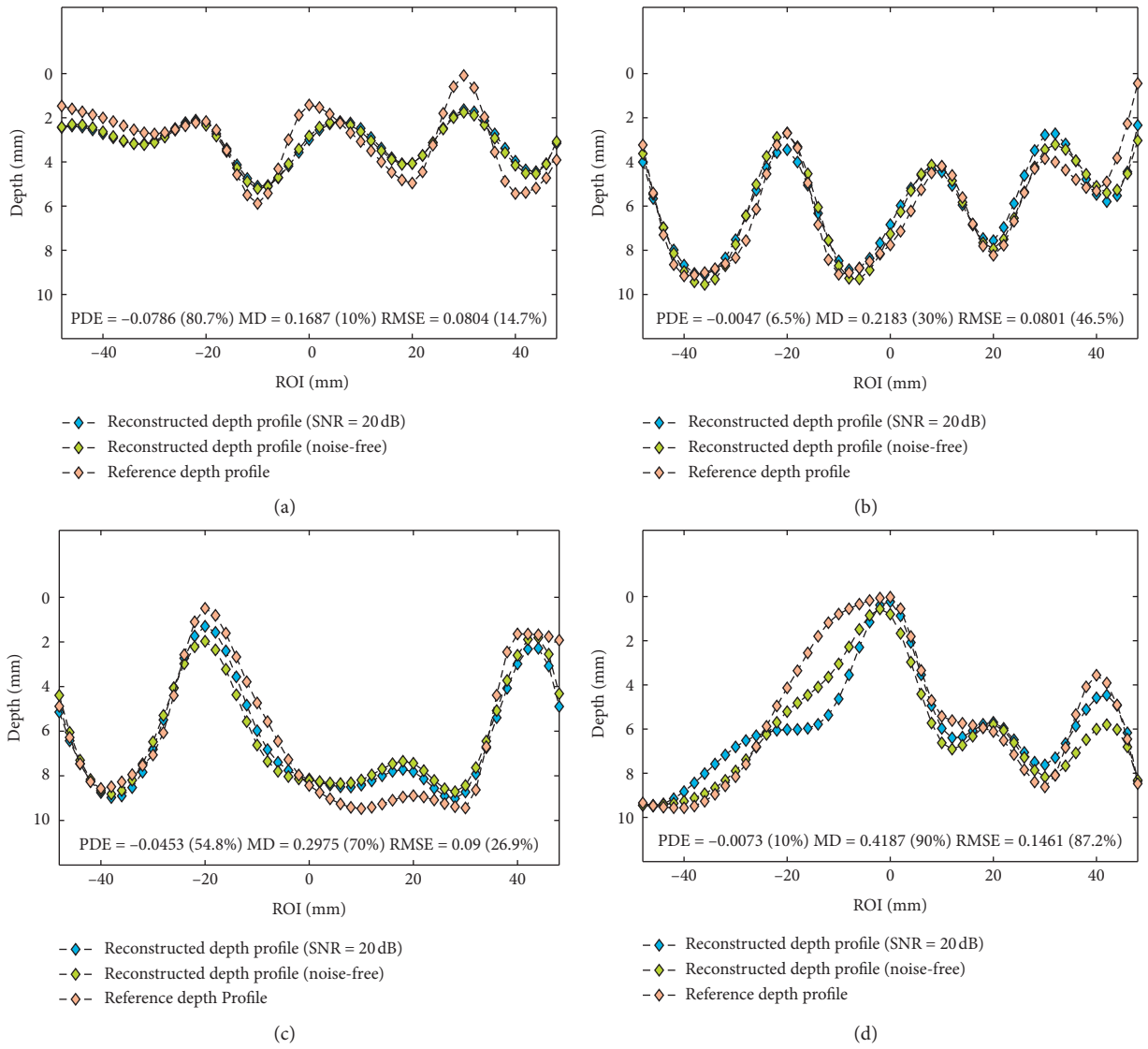


Figure 10: Sizing results with different ranking of MD value. (a) 10% ranking. (b) 30% ranking. (c) 70% ranking. (d) 90% ranking.

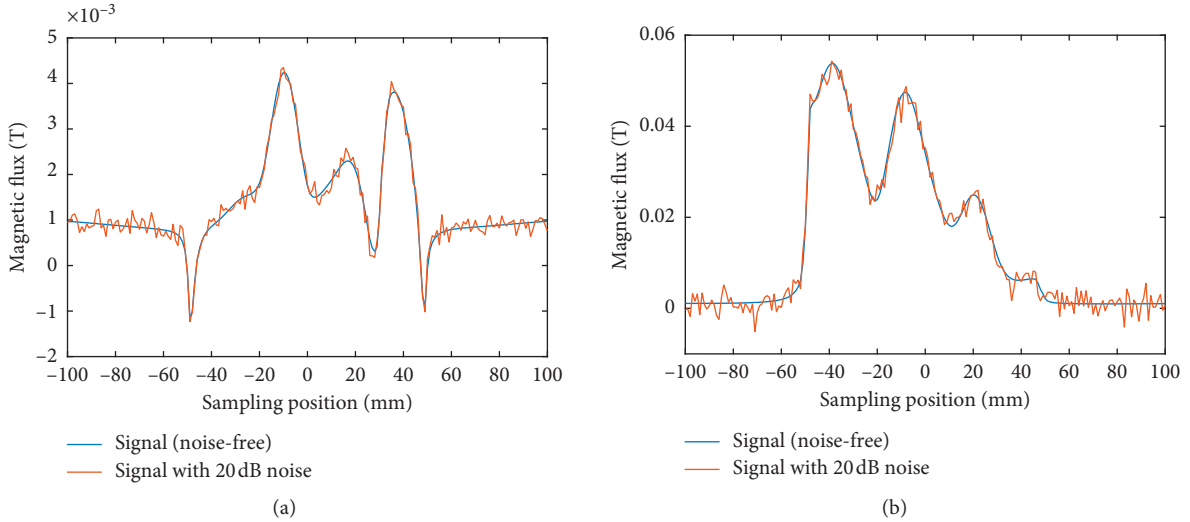


Figure 11: Continued.

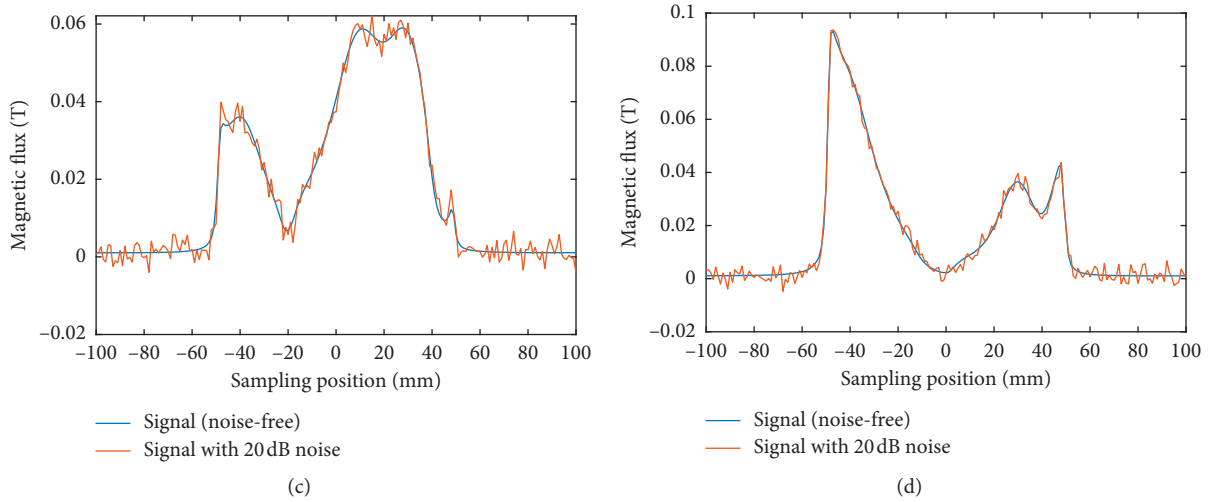


Figure 11: Signal used to get the results in Figure 10, including the contaminated signal and corresponding noise-free signal. (a) 10% ranking. (b) 30% ranking. (c) 70% ranking. (d) 90% ranking.

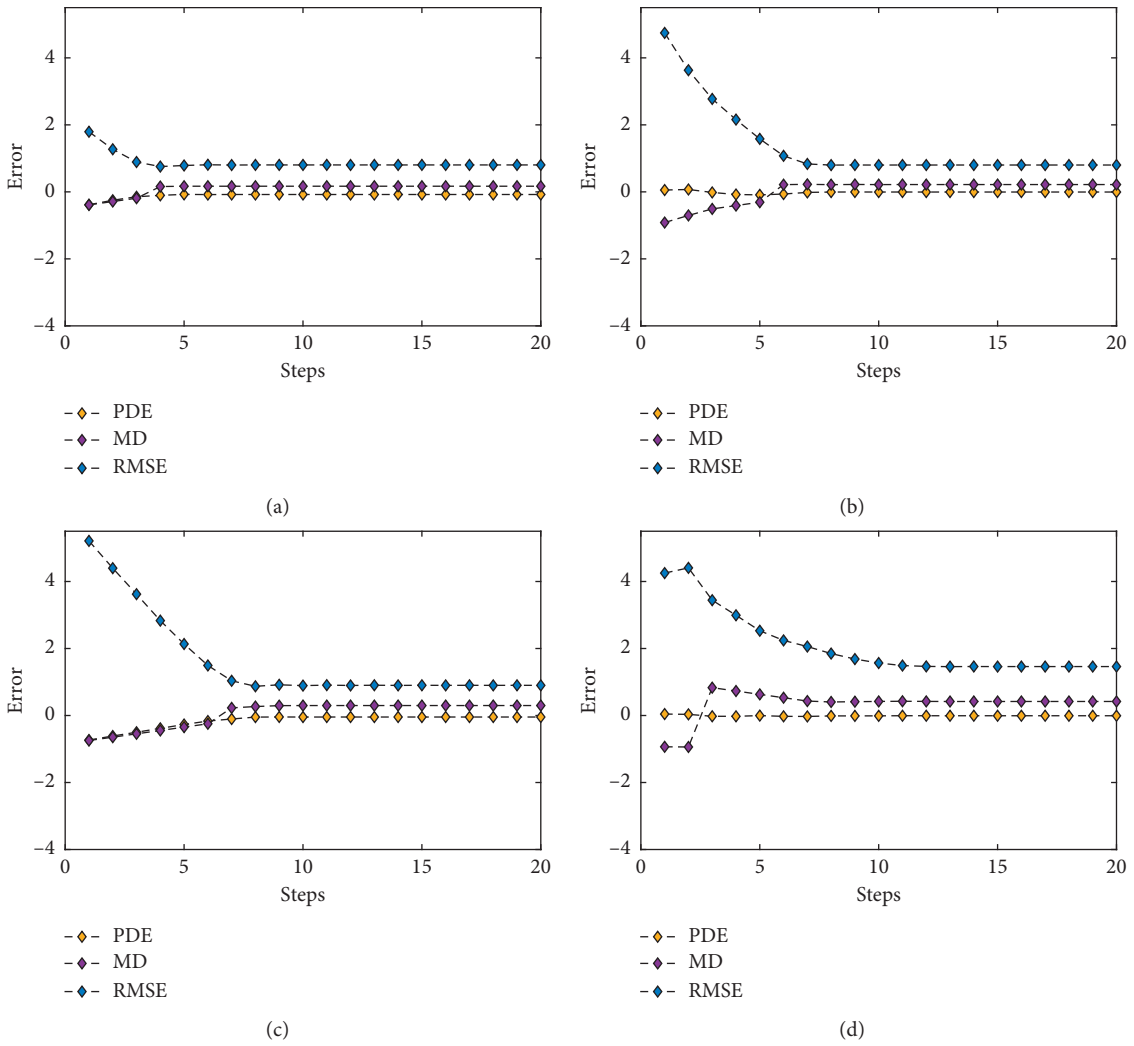


Figure 12: Error convergence for corresponding results in Figure 10. (a) 10% ranking. (b) 30% ranking. (c) 70% ranking. (d) 90% ranking.

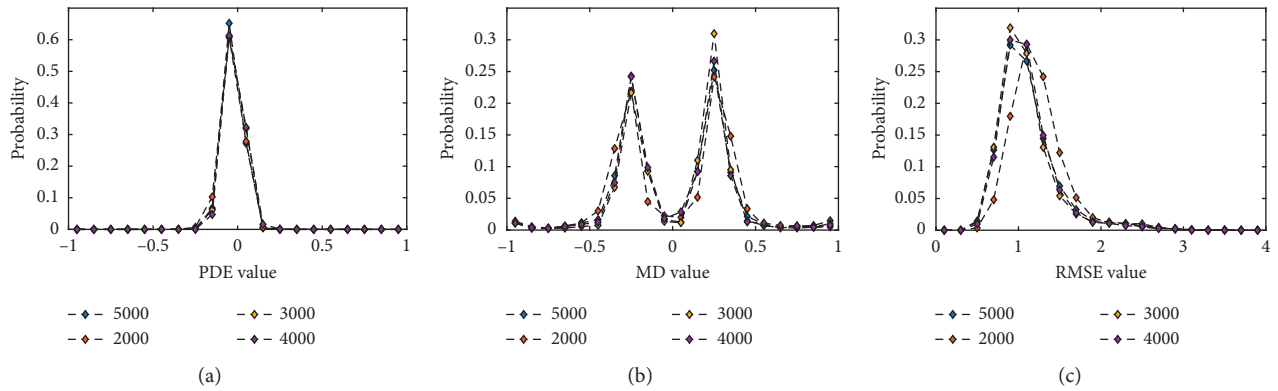


Figure 13: Error distribution of results from different size of training data sets. (a) PDE. (b) MD. (c) RMSE.

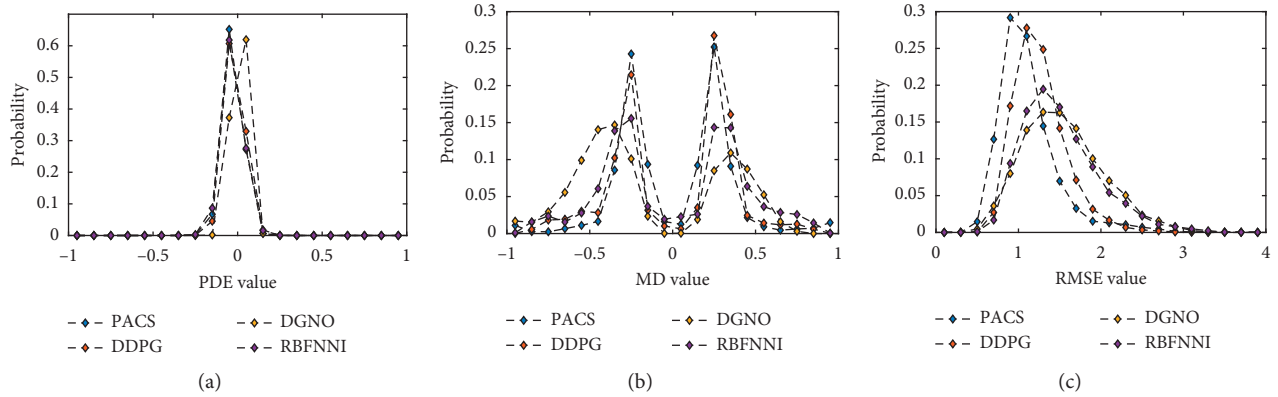


Figure 14: Results shown as distribution of error definitions from different reconstruction methods. (a) PDE. (b) MD. (c) RMSE.

From Figures 13(b) and 13(c), the results from 3000 to 4000 are even a little bit better than results from 5000. The results from 2000 have similar error distributions too, but it can also be seen that the performance is obviously not as good as results from larger training data sets from all aspects. The results from Figure 13 show that the quantity requirement of training data is not high and the algorithm is robust.

The actor-critic structured DDPG, the direct Gauss–Newton optimization (DGNO) in [14], and RBF neural network based iteration (RBFNNI) in [21] are selected as representative methods to show the accuracy and robustness of the algorithm proposed. The results are shown in Figure 14 in the form of error distribution too. The means of markers are the same as those in Figure 13. From Figures 14(b) and 14(c), it can be seen that the reconstruction results from PACS have obviously the best performance that the error distributions are more concentrated and closer to zero. From Figure 14(a), results from DDPG and DGNO are slightly better than those from PACS. Considering their performance on MD and RMSE, the results from PACS are still better than those from these methods.

5. Conclusion

In this paper, a pinning actor-critic structure-based solution for sizing complex depth profile with high degree of freedom of MFL inspection is studied. By involving the actor-critic structure, a novel way of utilizing the fine numerical based forward model in reconstructing the depth profile for MFL inspection is given. To solve the problem of the performance of reward deficiency, which is measured as Euclidean distance, a pinning strategy is given. By introducing the pinning subdefects, the action space has less variability than giving every subdefect an action. The robustness of the reconstruction results is improved by involving PACS.

The effectiveness of PACS proposed in this paper is tested with simulation results from nonlinear numerical forward models of MFL inspection with FEM. The results that are shown in a statistic way show the effectiveness of PACS proposed in this paper. The depth profiles reconstructed from signal with 20 dB noise are close to depth profiles reconstructed from noise-free signal, proving the robustness of PACS proposed. The results also show good accuracy compared with representative solutions of depth profile reconstruction.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 61703087 and 71502029).

References

[1] R. Palmer-Jones, S. Turner, and P. Hopkins, "A new approach to risk based pipeline integrity management," *International Pipeline Conference*, vol. 426, pp. 811–823, 2006.

[2] V. E. Loskutov, A. F. Matvienko, B. V. Patramanskii, and V. E. Shcherbinin, "The magnetic method for in-tube nondestructive testing of gas and oil pipelines: the past and the present," *Russian Journal of Nondestructive Testing*, vol. 42, pp. 493–504, 2006.

[3] Y. Shi, C. Zhang, R. Li, M. Cai, and G. Jia, "Theory and application of magnetic flux leakage pipeline detection," *Sensors*, vol. 15, no. 12, pp. 31036–31055, 2015.

[4] M. R. Kandroodi, B. N. Araabi, M. M. Bassiri, and M. N. Ahmadabadi, "Estimation of depth and length of defects from magnetic flux leakage measurements: verification with simulations, experiments, and pigging data," *IEEE Transactions on Magnetics*, vol. 53, no. 3, pp. 1–10, 2016.

[5] A. Joshi, L. Udpa, S. Udpa, and A. Tamburrino, "Adaptive wavelets for characterizing magnetic flux leakage signals from pipeline inspection," *IEEE Transactions on Magnetics*, vol. 42, no. 10, pp. 3168–3170, 2006.

[6] D. Mukherjee, S. Saha, and S. Mukhopadhyay, "Inverse mapping of magnetic flux leakage signal for defect characterization," *NDT & E International*, vol. 54, pp. 198–208, 2013.

[7] S. Saha, S. Mukhopadhyay, U. Mahapatra, S. Bhattacharya, and G. P. Srivastava, "Empirical structure for characterizing metal loss defects from radial magnetic flux leakage signal," *Ndt & E International*, vol. 43, no. 6, pp. 507–512, 2010.

[8] M. Hui and G. S. Park, "A study on the estimation of the shapes of axially oriented cracks in cmfl type ndt system," *IEEE Transactions on Magnetics*, vol. 50, no. 2, pp. 109–112, 2014.

[9] Z. Ma and H. Liu, "Pipeline defect detection and sizing based on mfl data using immune rbf neural networks," in *Proceedings of the 2007 IEEE Congress on Evolutionary Computation*, pp. 3399–3403, IEEE, Singapore, September 2007.

[10] Y. Li, J. Wilson, and G. Y. Tian, "Experiment and simulation study of 3d magnetic field sensing for magnetic flux leakage defect characterisation," *NDt & e international*, vol. 40, no. 2, pp. 179–184, 2007.

[11] W. Cheng, "Nondestructive testing of back-side local wall-thinning by means of low strength magnetization and highly sensitive magneto-impedance sensors," *IEEE Sensors Journal*, vol. 16, no. 14, pp. 5548–5556, 2016.

[12] S. Lu, J. Feng, F. Li, and J. Liu, "Precise inversion for the reconstruction of arbitrary defect profiles considering velocity effect in magnetic flux leakage testing," *IEEE Transactions on Magnetics*, vol. 53, no. 4, 1–12, 2016.

[13] M. Ravan, R. K. Amineh, S. Koziel, K. Natalia, and J. P. Reilly, "Sizing of 3-d arbitrary defects using magnetic flux leakage measurements," *IEEE Transactions on Magnetics*, vol. 46, no. 4, pp. 1024–1033, 2009.

[14] R. H. Priewald, C. Magele, P. D. Ledger, N. R. Pearson, and J. S. D. Mason, "Fast magnetic flux leakage signal inversion for the reconstruction of arbitrary defect profiles in steel using finite elements," *IEEE Transactions on Magnetics*, vol. 49, no. 1, pp. 506–516, 2012.

[15] K. C. Hari, M. Nabi, and S. V. Kulkarni, "Improved fem model for defect-shape construction from mfl signal by using genetic algorithm," *IET Science, Measurement & Technology*, vol. 1, no. 4, pp. 196–200, 2007.

[16] J. Feng, F. Li, S. Lu, and J. Liu, "Fast reconstruction of defect profiles from magnetic flux leakage measurements using a rbfnn based error adjustment methodology," *IET Science, Measurement & Technology*, vol. 11, no. 3, pp. 262–269, 2016.

[17] R. Schifini and A. C. Bruno, "Experimental verification of a finite element model used in a magnetic flux leakage inverse problem," *Journal of Physics D: Applied Physics*, vol. 38, no. 12, p. 1875, 2005.

[18] C. Edwards and S. B. Palmer, "The magnetic leakage field of surface-breaking cracks," *Journal of Physics D: Applied Physics*, vol. 19, no. 4, p. 657, 1986.

[19] Y. Zhang, Z. Ye, and C. Wang, "A fast method for rectangular crack sizes reconstruction in magnetic flux leakage testing," *Ndt & E International*, vol. 42, no. 5, pp. 369–375, 2009.

[20] X. Liu, Y. Deng, Z. Zeng, L. Udpa, and S. S. Udpa, "Model-based inversion technique using element-free galerkin method and state space search," *IEEE Transactions on Magnetics*, vol. 45, no. 3, pp. 1486–1489, 2009.

[21] J. Chen, S. Huang, and W. Zhao, "Three-dimensional defect inversion from magnetic flux leakage signals using iterative neural network," *IET Science, Measurement & Technology*, vol. 9, no. 4, pp. 418–426, 2015.

[22] W. Han, J. Xu, P. Wang, and G. Tian, "Defect profile estimation from magnetic flux leakage signal via efficient managing particle swarm optimization," *Sensors*, vol. 14, no. 6, pp. 10361–10380, 2014.

[23] F. Li, J. Feng, H. Zhang et al., "Quick reconstruction of arbitrary pipeline defect profiles from mfl measurements employing modified harmony search algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 9, pp. 2200–2213, 2018.

[24] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Playing atari with deep reinforcement learning," 2013, http://arxiv.org/abs/1312.5602.

[25] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[27] R. Wang, Q. Sun, W. Hu et al., "Soc-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, pp. 7866–7879, 2021.

[28] R. Wang, Q. Sun, D. Ma, and Z. Liu, "The small-signal stability analysis of the droop-controlled converter in electromagnetic timescale," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1459–1469, 2019.

[29] W. Rui, S. Qiuye, M. Dazhong, and H. Xuguang, "Line impedance cooperative stability region identification method for grid-tied inverters under weak grids," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 2856–2866, 2020.

[30] C. R. Taylor, *Applications of Dynamic Programming to Agricultural Decision Problems*, CRC Press, Boca Raton, FL, USA, 2019.

[31] V. T. Gerard, "A problem of dimensionality: a simple example," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 3, pp. 306-307, 1979.

[32] B. Chandrasekaran and A. K. Jain, "Quantization complexity and independent measurements," *IEEE Transactions on Computers*, vol. 100, no. 1, pp. 102–106, 1974.
[33] S. M. Dutta, F. H. Ghorbel, and R. K. Stanley, "Dipole modeling of magnetic flux leakage," *IEEE Transactions on Magnetics*, vol. 45, no. 4, pp. 1959–1965, 2009.
[34] D. A. G. Trevino, S. M. Dutta, F. H. Ghorbel, and M. Karkoub, "An improved dipole model of 3-d magnetic flux leakage," *IEEE Transactions on Magnetics*, vol. 52, no. 12, pp. 1–7, 2015.
[35] T. P. Lillicrap, J. J. Hunt, P. Alexander et al., "Continuous control with deep reinforcement learning," 2015, http://arxiv.org/abs/1509.02971.
[36] D. P. Kingma and B. Jimmy, "Adam: a method for stochastic optimization," 2014, http://arxiv.org/abs/1412.6980.
[37] D. Silver, L. Guy, N. Heess et al., "Deterministic policy gradient algorithms," in *Proceedings of the International Conference on Machine Learning*, pp. 387–395, PMLR, New York, NY, USA, July 2014.

Hindawi
Complexity
Volume 2021, Article ID 5584600, 16 pages
https://doi.org/10.1155/2021/5584600

Research Article

Joint Optimization of a Dry Port with Multilevel Location and Container Transportation: The Case of Northeast China

Feng Pian,[1] **Qiuju Shi**,[1] **Xue Yao**,[1] **Huiling Zhu**,[2] **and Weixin Luan**[1]

[1]*School of Maritime Economics and Management, Dalian Maritime University, Dalian 116026, China*
[2]*College of Transportation Engineering, Dalian Maritime University, Dalian 116026, China*

Correspondence should be addressed to Weixin Luan; weixinl@dlmu.edu.cn

Received 11 February 2021; Revised 17 March 2021; Accepted 31 March 2021; Published 22 April 2021

Academic Editor: Qiuye Sun

Dry port construction can reduce the cost of container transportation, and its location is the focus of existing research. Considering dry port capacity limitations and scale advantages, this study calculates the costs associated with dry port construction and operations, transportation, time, and the environment and constructs a joint optimization model of the dry port location and transportation scheme to minimize the total cost. Taking 35 prefecture-level cities in Northeast China as the source of container goods and Dalian port as the destination, this study conducts an empirical analysis using the Gurobi 9.0.2 optimizer of the AMPL software to solve the problem and takes the minimum total cost as the goal to select the best dry port and container transshipment scheme. The research draws the following conclusions. Seven dry ports also need to be built in the road-rail (RD-RL) mode, which shares 82.76% of the container transshipment volume, to reduce the total transportation cost by approximately 21.67%. Although multimodal transport through dry ports increases the time cost slightly, it can significantly reduce the economic and environmental costs of container transportation.

1. Introduction

In the process of economic and social development, the deployment and optimization of Cyber-Physical Systems play an increasingly important role. Power system, pipeline network, and transportation network are important components of Cyber-Physical Systems. Some achievements of energy system and pipeline network [1–3] provide a new idea for the study of connectivity reliability of urban transportation network. The joint optimization model proposed in this study solves the optimization problem of transportation network, which can be used as reference for other physical network optimization such as power system or pipe network. Similar to other networks, inland container transportation network has the structural characteristics of a complex network. However, the spatial entity of transportation network makes it different from an abstract network such as the social network, which is especially apparent in inland container transportation network. Due to the uncertainty of construction conditions, grade of dry port, freight demand, and transshipment scheme, the joint optimization of dry port selection and container transshipment presents additional complexity.

Dry ports can optimize the inland transportation system by reducing container transportation costs, alleviating port congestion, expanding the hinterland of the seaport, and providing convenient services for shippers (such as customs declaration) [4, 5]. Moreover, some studies have confirmed that dry ports have significant effects in reducing carbon dioxide emissions, transportation costs, and waiting time for container trucks [6–8]. Dry ports can also effectively integrate various inland transportation methods and nodes, improve the efficiency and benefits of inland cargo transportation, and promote the rapid development of inland container multimodal transportation [9]. With rapid globalization and developments in international trade, the container throughput of China's coastal ports has greatly expanded, reaching 261.07 million TEU in 2019 [10]. Reductions of cost, energy consumption, and carbon emissions of inland container transportation contribute to the

sustainable development of the Chinese economy, especially to the logistics industry [11]. In addition, there is a negative decoupling between China's economic development and freight demand, and there is a limit to the growth of freight demand [12]. With the growth of gross domestic product (GDP), freight demand remains stable or even declines [12], and China's seaports are bound to face increasing competition for supply in the future.

As a hub in the transportation network, the role and value of the dry port depend on its location advantages. A reasonably located dry port can attract enough goods in the inland areas and achieve economies of scale through a combination of rail and road transport. On the contrary, problems such as overcapacity, waste of resources, low efficiency and utilization, and low rate of return on investment may occur in poorly planned dry ports. Moreover, the construction of a dry port requires a large amount of capital investment; once it is completed, it can hardly be relocated, and the sunk cost is high [13]. More than 150 dry ports were built in China by December 2019 (according to the data collected from Statistics). Most of them are not connected by railways, and some even lack the function of customs clearance. The throughput of goods is not consistent with the input of resources, and the efficiency of these dry ports is low [14]. As an important logistics node in the inland container transportation network, the location layout, scale, quantity, and service area of the dry port have important contributions to the efficiency of the entire transportation chain. This study measures the economic and environmental benefits of developing a dry port, considers the dry port's carrying capacity, sets up its construction and operational costs at different scale levels, quantifies the time and carbon emission costs, builds the location model of the dry port, seeks the optimal scheme of inland container transportation, and takes Northeast China as an example. Based on the freight distribution of the foreign trade container volume, the optimal number of dry ports, location layout, service scope, and transportation scheme of the region are obtained to prove the effectiveness of the model.

2. Literature Review

2.1. Optimization of Container Transportation

2.1.1. Optimization of Transportation Mode, Route, and Node. The main goal of related research is cost minimization, and the research methods mainly include (multiple) goal, mixed integer, and bi-level programming models. Jeong et al. [15] established a linear integer programming model with objective functions, including operation and transportation time costs, to reveal the potential hub location, and Wang and Yun [16] studied the container transportation problem with time windows under truck and train transportation modes. Demir et al. [17] proposed a green multimodal transport service network design problem with uncertain freight transport time window and demand and, using the sample average approximation method, provided a flexible, reliable, and environmentally friendly alternative for long-distance transportation of large quantities of goods. Wang et al. [18] studied the modeling and optimization of a combined highway-railway transportation system embedded in a hub spoke network under uncertainty and proposed a fuzzy biobjective optimization method that minimizes the expected value of the total cost and the maximum time demand on the critical value, so as to optimize the combined highway-railway transportation mode. Zhao et al. [19] established a two-stage model of combined positioning and path planning to locate a river hub port and dry port with the objective of minimizing total transportation cost.

2.1.2. Optimization of a Dry Port-Seaport Network. Most of the related research considers carriers, seaports, dry port operators, and cargo owners and integrates economic, social, and environmental factors to maximize the interests of all parties, so as to optimize the dry port-seaport network. Wang et al. [20] studied the location of a dry port considering the interaction between the port and hinterland, comprehensively considered the relationship among the dry port, seaport, and regional logistics system, and optimized the configuration of the dry port system. Chang et al. [21] considered the carrying capacity limitation of the dry port, constructed a two-stage multicapacity-level dry port location model with the objective of minimizing construction and transportation costs, and optimized the layout and freight demand distribution of the dry port in Northeast China. The game theory method is also introduced into the optimization of dry port–seaport network [22–25]. Wei and Dong [26] applied the biobjective, mixed integer programming model to study a new type of cross-border logistics network connecting the marine logistics network with the inland cross-border logistics network through the dry port, and discussed the organization optimization of inland import and export goods under different network scenarios. Tsao and Thanh [27] used a multiobjective hybrid robust possibilistic flexible programming method to determine the optimal number, location, and capacity of dry ports. Van Nguyen et al. [13] combined data mining and complex network theory and used a two-stage optimization method to determine the location and service area of a dry port in a large-scale inland transportation system.

2.2. Cost Accounting of Container Inland Transportation. In recent years, the concept of sustainable development has been deeply rooted in the hearts of people, and transportation demand has become increasingly diversified. Environment, time, and social costs are becoming increasingly important in the optimization of inland container transportation [28]. As the dry port becomes the key hub of the inland container transportation network, the construction and operation costs of the dry port get included in the comprehensive cost of inland container transportation. O'Kelly [29] proposed a hub location model considering the fixed cost of infrastructure, added the fixed cost to the single allocation hub location problem, and optimized the p-hub location model with the number of hubs as the decision variable. Janic [30] proposed a comparable internal and external comprehensive cost model, and discussed the

influence of the policy of internalizing external transport costs on the future competition between intermodal transport and road freight networks. Zhang et al. [31] internalized CO_2 emission cost and optimized the network configuration for different CO_2 prices. Chang [32] established a total cost optimization model including fixed and transportation costs. Pekin et al. [33] considered the impact of cargo types on time value, established a total cost function including multimodal transport market price and container time value, and improved the multimodal transport terminal location analysis model. Rahimi et al. [34] developed a multiobjective, multimodal queuing system considering traffic congestion and the maximum carrying capacity of the hub to analyze the waiting time of flow units in a transportation hub; this system was used to solve the multiobjective hub location problem of hub congestion. Wei and Dong [26] used comprehensive costs, including economic, environmental, and social costs, to study the long-term sustainable development of the dry port network.

Previous studies on the locations of dry ports have many achievements (Table 1). Based on previous studies, this study uses a technical and economic method to convert the construction cost of the dry port by year, brings the operation and time costs of the dry port into the comprehensive cost, optimizes the calculation of container road transportation cost, and considers the scale effect of railway transportation, as well as the capacity level of the dry port and the conditions of urban construction of the dry port, so as to make the joint optimization model of the dry port location and transportation scheme closer to the actual problems. Then, taking China's northeast region as the empirical object, the empirical calculation is carried out, and the solution is solved using the AMPL software Gurobi 9.0.2 optimizer. The cargo flow distribution of container cargo is examined with regard to when it can be split or not under three transport modes: road transport, intermodal road-rail-road transport, and intermodal road-rail transport; the total cost of container transport under different transport scenarios are discussed, and the scale, location layout, and service scope of dry port construction when the total cost is optimal are determined, so as to realize the optimization of dry port location and inland container transport.

3. Problem Description

The joint optimization problem of the dry port location and transportation scheme studied in this paper is a two-stage and three-level inland container transportation network (Figure 1) optimization problem, which is composed of shippers, dry ports, and seaports. In this network, there are three modes of transportation: the first mode is intermodal road-rail transport (RD-RL mode), in which containers are transported from the shipper to the dry port by road, and then transported from the dry port to the seaport by rail. The second mode is intermodal road-rail-road transport (RD-RL-RD mode). In reality, some dry ports lack railway facilities that are directly connected with the seaport and thus need road trucks to transport containers from the seaport city railway station to the seaport wharf. The third mode is road transport (RD mode), in which the containers are directly transported from the shipper to the seaport.

For the inland container transportation network, Net = (Nd, E), where Netrefers to the inland container transportation network and $\mathrm{Nd} = \{\mathrm{Nd}^1, \mathrm{Nd}^2, \mathrm{Nd}^3\}$ represents the transportation node. There are three types of nodes in the network: $\mathrm{Nd}^1 = \{\mathrm{Nd}_i^1, i = 1, 2, 3, \ldots, I\}$ refers to the node of the shipper city$I$; $\mathrm{Nd}^2 \subseteq \mathrm{Nd}^1 = \{\mathrm{Nd}_j^2, j = 1, 2, \ldots, J\}$ refers to the node of the dry port city J; and the dry port node is generated in the shipper city node. $\mathrm{Nd}^3 = \{\mathrm{Nd}_k^3, k = 1\}$ represents the only seaport node, and $E$is the arc between the nodes Nd^1, Nd^2, and Nd^3, which are different modes of transportation. $E_{ijk} = \{\mathrm{Nd}_i^1, \mathrm{Nd}_j^2, \mathrm{Nd}_k^3\}$ is the arc that refers to road rail transportation through the dry port. $E_{ik} = \{\mathrm{Nd}_i^1, \mathrm{Nd}_k^3\}$ is the arc that means the road is directly transported.

4. Formulation

4.1. Model Assumptions

(1) The total cost of the inland container transportation network includes the construction, operation, and logistics (transportation economic, time, and carbon emission) costs of the dry port.

(2) Under the corresponding construction level, the container transfer capacity limits, fixed construction cost, annual operation cost, recommended scale range, storage waiting time, and discount coefficient of the scale effect of rail tariff are known.

(3) The foreign trade volume of containers in node cities, namely, freight demand, is known (dry ports are generated in shipper city nodes).

(4) The containers are transferred once through a single dry port.

(5) Without considering the transport capacity constraints in the freight transportation process, sufficient road passing capacity and sufficient railway trains are assumed.

(6) During container transport, trucks and trains travel at a known average speed.

(7) The container transport in the inland container transport network is a one-way transport, and only the process of transporting containers from the shipper to the seaport is studied; reverse transportation is not considered.

(8) A 20-foot box is the object of study, and the total weight of the container is set at 24 tons according to the standard set by ISO/TC104.

(9) The maximum level limit of dry ports that can be built in different cities is known.

4.2. Model Formulation. The parameters and variables in cost accounting and model construction and their meanings are listed in Table 2.

Taking the minimum total cost of inland container transportation as the objective function, this study

Table 1: Summary of the literature on dry port location problem.

	Global perspective	Grade constraint	Factors to be considered				Construction cost	Road cost function		Problem size		Method
			Economy	Environment	Social	Time		Linear	Subsection	Medium	Large	
Feng et al.(2013)			√				√	√		√		Greedy algorithm; genetic algorithm
Chang Zheng et al (2017)		√	√				√	√		√		MCHLP; genetic algorithm
Zhang et al.(2018)			√			√	√	√		√		Game-theoretical; logit discrete selection
Tsao and Linh (2018)	√		√	√	√		√	√		√		Continuous approximation; game theory
Xu et al.(2018)			√	√		√		√		√		Game-theoretical; logit discrete selection
Tsao and Thanh (2019)		√	√	√	√		√	√		√		MOMRPFP
Zhao et al.(2019)	√		√			√		√		√		Two-stage planning model
Van Nguyen et al.(2020)	√										√	Data mining; complex network
This paper	√	√	√	√		√	√		√	√		MCHLP; AMPL

"Global perspective" does not specify the alternative dry port cities, but directly determines the location of dry port from the city nodes. "Problem size" refers to the size of the studied dry port network, which is classified as medium-scale if the study focuses on regional-level network and as large-scale if the focus is on the nationwide network.

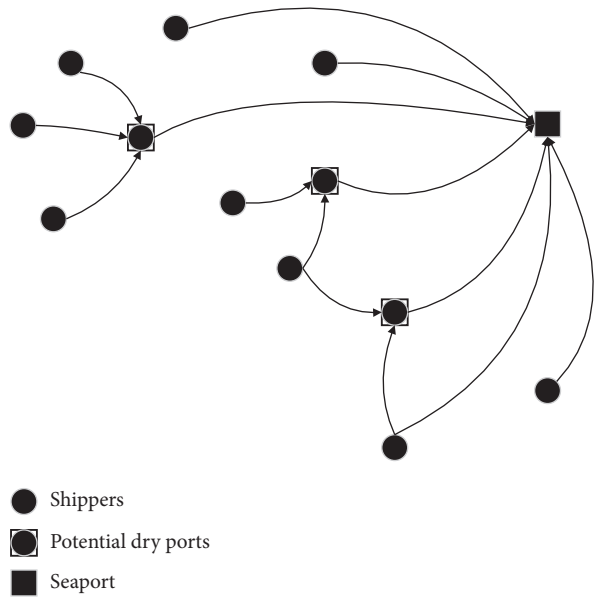


Figure 1: A simple diagram of an inland container transportation network.

Table 2: The parameters and variables and their meanings.

Symbol	Category	Meaning
I	Indices	Set of shippers
J	Indices	Set of candidate locations of dry ports
G	Indices	Set of capacity levels for dry ports
f_b	Parameters	The total cost of opening a dry port with capacity level b (USD)
s	Parameters	Investment payoff period for dry port construction (year)
u	Parameters	Capital discount rate (percentage)
r_g	Parameters	The cost of operating a dry port with capacity level g per year (USD)
C_{ij}^{rd}	Parameters	Economic cost of container transport from shipper i to dry port j by road (USD)
C_{jk}^{rl}	Parameters	Economic cost of container transport from dry port j to seaport k by railway (USD)
C_{kk}^{rd}	Parameters	Economic cost of container transportation from the rail station in the seaport city to seaport k by road (USD)
L_{ij}^{rd}	Parameters	Distance of transport from shipper i to dry port j by road (km)
L_{jk}^{rl}	Parameters	Distance of transport from dry port j to seaport k by rail (km)
L_{kk}^{rd}	Parameters	Distance of container transport from the railway station in the seaport city to seaport k by road (km)
P^1	Parameters	Base price of railway delivery (USD)
P^2	Parameters	Base price of railway operation (USD)
P^3	Parameters	Container usage fee (USD)
P^4	Parameters	Container cleaning fee (USD)
P^5	Parameters	Container handling charges (USD)
P^6	Parameters	Unloading charges of containers in the yard (USD)
disc_g	Parameters	Discount coefficient of the dry port with capacity level g
W	Parameters	The price of transport per container from the rail station in the seaport city to seaport s by road (USD per TEU)
L_{ik}^{rd}	Parameters	Distance of transport from shipper g to seaport s by road (km)
C_{ij}^{rdt}	Parameters	Time cost of container transport from shipper i to dry port j by road (USD)
C_{jk}^{rlt}	Parameters	Time cost of container transport from dry port j to seaport k by rail (USD)
C_{kk}^{rdt}	Parameters	Time cost of container transport from the rail station in the seaport city to seaport k by road (USD)
t_g	Parameters	Cargo storage time at the dry port with capacity level g (day)
V	Parameters	Average value of the cargo per container (USD)
R	Parameters	Current deposit interest rate (percentage)
v^{rl}	Parameters	Average speed of container transport by rail (km per hour)
v^{rd}	Parameters	Average speed of container transport by road (km per hour)
DN	Parameters	365, the number of days in a year
HN	Parameters	24, the number of hours in a day
C_{ij}^{rdc}	Parameters	Carbon dioxide emission cost of container transport from shipper i to dry port j by road (USD)
C_{jk}^{rlc}	Parameters	Carbon dioxide emission cost of container transportation from dry port j to seaport k by railway (USD)

Table 2: Continued.

Symbol	Category	Meaning
C_{kk}^{rdc}	Parameters	Carbon dioxide emission cost of container transport from the railway station in the seaport city to seaport s by road (USD)
E^{rl}	Parameters	Carbon dioxide emissions per container per distance by rail (kg eq per container per km)
E^{rd}	Parameters	Carbon dioxide emissions per container per distance by road (kg eq per container per km)
E^{rr}	Parameters	Carbon dioxide emissions per container of road and rail transshipment handling (kg eq per TEU)
T^c	Parameters	Unit cost of carbon dioxide emissions handling in the outside market (USD per kg eq)
n	Parameters	$\begin{cases} 2, & \text{RD} - \text{RL} - \text{RD mode}, \\ 1, & \text{RD} - \text{RL mode}. \end{cases}$
Cap_g	Parameters	Capacity of the dry port with capacity g (TEU)
H_j	Parameters	The highest level of the dry port that the candidate dry port city j can build
M	Parameters	1,000,000,000, a large enough number
Q_i	Parameters	Volume of containers transported from shippers $i \in I$
y_{jg}	Decision variables	$\begin{cases} 1, & \text{if dry port } j \text{ with capacity level } g \text{ is opened}, \\ 0, & \text{otherwise}. \end{cases}$
x_{ij}	Decision variables	Whether containers are transported from shippers $i \in I$to candidate dry ports $j \in J$
q_{ij}	Decision variables	Volume of containers transported from shippers $i \in I$to candidate dry ports $j \in J$

establishes a model of a dry port location with multiple capacity levels. The total transportation cost, C, is composed of the annual dry port construction cost C^c, dry port operating cost per year C^o, and comprehensive transportation cost C^{lc}, which can be expressed as

$$C = C^c + C^o + C^{lc}. \tag{1}$$

4.2.1. Annual Construction Cost of Dry Port C^c. Assuming that the present value of investment is f, the benchmark rate of return on investment is$u$, and the construction calculation period of the project is $s$years, then the capital recovery annuity is$f \times u/[1 - (1+u)^{-s}]$, and $f \times u/[1 - (1+u)^{-s}]$is the annual capital recovery coefficient [35], from which the annual construction cost of grade g dry port can be obtained:

$$C^c = \sum_{j\in J}\sum_{g\in G} f_g y_{jg} \frac{u(1+u)^s}{(1+u)^s - 1}. \tag{2}$$

4.2.2. Operating Cost of Dry Port C^o. The operating cost of the dry port refers to the costs directly related to the operation of the dry port or related to the labor services provided by the dry port, including direct material expenses, direct labor costs, and indirect costs of operation and management. Transshipment is the main service provided by dry ports. In this study, the cost of transshipment between road and rail transport that occurs in dry port transshipment is classified as part of the operating cost of dry ports.

$$C^o = \sum_{j\in J}\sum_{g\in G} r_g y_{jg}. \tag{3}$$

4.2.3. Comprehensive Transportation Costs of Inland Containers C^{lc}. The comprehensive transportation cost of inland containers comprises transportation, time, and environmental costs.

This study assumes that the scale effect of a dry port is reflected in the economic cost of railway transportation. The larger the level of the dry port, the smaller is the discount coefficient of the scale effect. The market-oriented degree of container road transportation is high; thus, there is no clear formula to determine its price. Assuming that the container freight from the railway station to the seaport in the seaport city is constant (W), and the rest of the road freight is a function of the road transportation distance $f(L)$, the specific functional relationship needs to be calculated by collecting the actual data. The economic cost of container transport through dry port transshipment is

$$\begin{aligned} C^{dp} &= \sum_{i\in I}\sum_{j\in J} C_{ij}^{rd} + \sum_{j\in J}\sum_{i\in I}\sum_{g\in G} C_{ij}^{rl} + (n-1)\sum_{i\in I}\sum_{j\in J} C_{kk}^{rd} \\ &= \sum_{i\in I}\sum_{j\in J} f\left(L_{ij}^{rd}\right) q_{ij} + \sum_{j\in J}\sum_{i\in I}\sum_{g\in G}\left[\left(P^1 + P^2 L_{jk}^{rl} + P^3 + P^4 + P^5 + P^6\right) q_{ij}\right] \text{disc}_g + (n-1)\sum_{i\in I}\sum_{j\in J} W q_{ij}. \end{aligned} \tag{4}$$

The economic cost of container transportation for direct road transport is

$$C^{rd} = \sum_{i\in I}\sum_{j\in J} f\left(L_{ik}^{rd}\right)\left(Q_i - q_{ij}\right). \tag{5}$$

In this study, the time cost of transportation is expressed using the current interest that can be generated by the value of the freight in the container during transportation for the corresponding time. The time cost of container transport through dry port transit is calculated as

$$C^{dpt} = \sum_{i\in I}\sum_{j\in J} C_{ij}^{rdt} + \sum_{j\in J}\sum_{i\in I}\sum_{g\in G} C_{jk}^{rlt} + \sum_{i\in I}\sum_{j\in J} C_{kk}^{rdt}$$
$$= \sum_{i\in I}\sum_{j\in J} \frac{L_{ij}^{rd}}{\mathrm{DN}\cdot v^{rd}\cdot \mathrm{HN}} \mathrm{VR}\, q_{ij} + \sum_{j\in J}\sum_{i\in I}\sum_{g\in G} \frac{1}{\mathrm{DN}} \left(\frac{L_{jk}^{rl}}{v^{rl}\cdot \mathrm{HN}} + t_g\right) \mathrm{VR}\, q_{ij} + \sum_{i\in I}\sum_{j\in J} \frac{L_{kk}^{rd}}{\mathrm{DN}\cdot v_{rd}\cdot \mathrm{HN}} \mathrm{VR}\, q_{ij}. \tag{6}$$

The container transportation time cost of direct road transportation is

$$C^{rdt} = \sum_{i\in I}\sum_{j\in J} \frac{L_{ik}^{rd}}{\mathrm{DN}\cdot v^{rd}\cdot \mathrm{HN}} VR\left(Q_i - q_{ij}\right). \tag{7}$$

In this study, the carbon tax to be paid for CO_2 emissions is used to express the environmental cost, which depends on the amount of CO_2 emissions during the transportation, loading, and unloading of containers. The carbon emission cost of container transportation by dry port transshipment is

$$C^{dpc} = \sum_{i\in I}\sum_{j\in J} C_{ij}^{rdc} + \sum_{j\in J}\sum_{i\in I} C_{jk}^{rlc} + C_{rdc}^{ss}$$
$$= \sum_{i\in I}\sum_{j\in J} E^{rd} L_{ij}^{rd} q_{ij} T^c + \sum_{j\in J}\sum_{i\in I}\left(E^{rl} L_{jk}^{rl} + nmq_{ij} T^c\right) + \sum_{i\in I}\sum_{j\in J} E^{rd} L_{kk}^{rd} q_{ij} T^c. \tag{8}$$

Cost accounting of the carbon emissions of container transportation for direct road transportation is

$$C^{rdc} = \sum_{i\in I}\sum_{j\in J} E^{rd} L_{ik}^{rd}\left(Q_i - q_{ij}\right) T^c. \tag{9}$$

In summary, the logistics-related costs of inland container transport are accounted for as

$$C^{lc} = C^{dp} + C^{rd} + C^{dpt} + C^{rdt} + C^{dpc} + C^{rdc}. \tag{10}$$

The model is as follows:

$$\begin{aligned}
&\min C = C^c + C^o + C^{lc},\\
&\text{s.t.}\\
&Q_i, q_{ij} \in N, \quad \forall i \in I, \forall j \in J,\\
&\sum_{j\in J} q_{ij} \le Q_i, \quad \forall i \in I,\\
&x_{ij} M \ge q_{ij}, \quad \forall i \in I, \forall j \in J,\\
&x_{ij} \le q_{ij} M, \quad \forall i \in I, \forall j \in J,\\
&\sum_{g\in G} y_{jg} \le \sum_{i\in I} x_{ij}, \quad \forall j \in J,\\
&M\sum_{g\in G} y_{jg} \ge \sum_{i\in I} x_{ij}, \quad \forall j \in J,\\
&\sum_{g\in G} y_{jg} \le 1, \quad \forall j \in J,\\
&\sum_{i\in I} q_{ij} \le \sum_{g\in G} y_{jg} \mathrm{Cap}_g, \quad \forall j \in J,\\
&\sum_{g\in G} y_{jg} g \le H_j, \quad \forall j \in J,\\
&y_{jg} \in \{0,1\}, \quad \forall j \in J, \forall g \in G.
\end{aligned} \tag{11}$$

5. Case Study of Northeast China

In 2017, Dalian Port undertook 98.5% of the volume of foreign trade containers in Northeast China. The sea-rail intermodal transport channel, with Dalian port as the main sea-rail intermodal hub, has a large-scale operation, which delivers the sea-rail intermodal transport volume throughout Northeast China. The foreign trade container hinterland of Dalian port includes three provinces in Northeast China and four leagues in Inner Mongolia. Due to the particularity of geographical location, the hinterland has strong independence. In addition, in 2016, among the major ports in China, Dalian Port's sea-rail combined transportation volume reached 406,000 TEU, accounting for 19.16% of the country's total volume. To sum up, it is representative and typical to select the northeast region to carry out the joint optimization research of dry port planning and container transportation.

5.1. Research Scope. There are 41 prefecture-level cities in Northeast China, including six coastal port cities. The container transportation modes of these cities are more diverse than those of inland cities. They are not included in the scope of this study. Therefore, the scope of shippers and alternative dry port cities in this study covers 35 prefecture-level cities in the inland of Northeast China (Table 3), and the destination port is Dalian Port.

5.2. Data Sources and Processing. The core of the joint optimization problem of dry port with multilevel location and container transportation is to determine the number, location, and grade of land ports, so as to achieve the goal of reducing inland container transportation cost. The total transportation cost is selected as the objective function. The basis of joint optimization is to determine the cost of road

Table 3: Thirty-five cities in Northeast China.

Research scope	City name
Heilongjiang province	Harbin, Qiqihar, Jixi, Hegang, Shuangyashan, Daqing, Yichun, Jiamusi, Qitaihe, Mudanjiang, Heihe, Suihua and Daxinganling
Jilin province	Changchun, Jilin, Siping, Liaoyuan, Tonghua, Baishan, Songyuan, Baicheng, and Yanbian
Liaoning province	Shenyang, Anshan, Fushun, Benxi, Fuxin, Liaoyang, Tieling, and Chaoyang
Inner mongolia autonomous region	Chifeng, Tongliao, Hulunbuir, Xing'an league, and Xilin Gol League

and rail transportation. The demand and spatial distribution of container freight also have an important influence on joint optimization. In addition, cities with developed economy and convenient transportation should be chosen for the construction of dry ports. On one hand, it is conducive to the support of goods and human resources; on the other hand, it can facilitate the distribution of goods and enhance the attraction of goods by dry port. Therefore, this study collects data regarding the container transportation volume of each city, the road distance and railway distance between cities, and according to the existing research, the average speed of road and rail transportation, and other related parameters are set.

5.2.1. Mileage Data. The railway distance and the highway distance (Table 4) were taken from the "Train ticket network" and the "Gaode map," respectively (the data for "Train ticket network" is updated up to 2020).

5.2.2. Foreign Trade Container Weights. From custom data statistics, this study obtained the volume of the foreign trade containers of 35 cities (Table 4) in Northeast China in 2017.

5.2.3. Transportation Cost-Related Data. Reflecting the actual situation in Northeast China, this study sets the average speed of highway operation as 70 km/h, and the average speed of railway operation as 100 km/h. The relevant charging standards of rail transport (Table 5) come from the railway freight tariff rules of the Ministry of Railways of the People's Republic of China and the notice of the National Development and Reform Commission on issues related to the adjustment of railway freight tariffs (2014).

The relevant fee standard for road transport is regressed from the cost data of 83 road containers (20-ft. containers) transported by 29 freight forwarders in China (Table 6) [36]. The price function of the container road transport is

$$f(L_R) = \begin{cases} 1204.31, & L_R \le 50, \\ 0.0007L_R^2 + 5.7532L_R + 914.9, & L_R > 50. \end{cases} \quad (12)$$

Among them, $L_R$is the road transport distance.

Regarding the economic cost of road transport from the railway station to the port in the seaport city when the seaport city lacks railway facilities that are directly connected to the port, the container goods must be transported to the port through the container truck. Generally, the distance between the railway station of the port city and the port is short, and it is easy to distribute goods back and forth; this is classified as short-haul transportation. Through investigation, it is concluded that the price (W) of short-distance transshipment of containers in the seaport city is 29.61 USD/TEU.

5.2.4. Time Cost-Related Data. According to the average value of export containers in Dalian port, $V = 29{,}609.01$ USD/TEU; the interest rate is the current deposit interest rate in 2020; and $R = 0.35\%$.

5.2.5. Carbon Cost Data. According to Li and Su [37], carbon emissions from road transport are quantified as 0.796 kg/(ton· km); that from rail transport, 0.028 kg/(ton· km); and that from rail transit, 1.56 kg/ton. According to the standard of the Ministry of Environmental Protection, the carbon tax is set at 7.40 USD/ton. According to the ISO Standard for Container, the weight of the 20-foot container is 24 tons.

5.2.6. Calculation Parameters of Land Port Construction Cost. In this study, the capital discount rate is 8%, and the capital recovery period for dry port construction cost is 15 years.

5.2.7. Relevant Data of Dry Port Grade. Referring to the relevant literature [38] and the investment and operation of similar dry ports in China, this study estimates the construction level and annual operation cost of the dry port, recommended scale range, general storage waiting time, and the discount coefficient of the scale effect and divides the dry port into four levels. The scale of the first-level dry port is the smallest, and the scale of the fourth-level dry port is the largest. The specific divisions are listed in Table 7.

5.3. The Capacity of Cities to Build Dry Ports. The construction of a dry port is closely related to the city's economic development level, investment capacity, foreign trade development level, transportation conditions, and railway infrastructure conditions. Therefore, this study selects nine indicators related to the construction and development of dry ports: GDP, actual foreign investment, total retail sales of social consumer goods, local government revenue, investment in fixed assets, total post and telecommunications business, foreign trade import, foreign trade export, and the score of railway station grade (the data corresponding to the index is from the city statistical bulletin in 2018). The factor analysis method is used to evaluate 35 inland cities in

Table 4: Foreign trade container volume and related mileage data of cities.

	Foreign trade container weight in 2017/TEU	Road distance to Dalian port/km	Railway distance to Dalian station/km
Anshan	54799	302.2	308
Liaoyang	25145	332.3	333
Shenyang	467953	382.1	397
Fushun	31920	445.9	444
Benxi	20486	381.2	457
Tieling	8408	452.7	467
Chaoyang	30800	481.3	473
Fuxin	6793	408.4	535
Liaoyuan	7254	611.0	559
Siping	9763	582.4	585
Tongliao	17664	633.8	634
Changchun	395951	680.8	700
Jilin	36360	782.7	821
Baishan	14326	637.3	833
Songyuan	9654	812.5	849
Tonghua	3960	557.4	870
Baicheng	15547	902.5	938
Harbin	45141	948.6	946
Chifeng	20761	633.5	968
Suihua	12569	1051.9	1071
Yanbian Korean autonomous prefecture	49019	1016.7	1073
Xing'an league	293	963.5	1116
Daqing	81033	985.3	1120
Qiqihar	5300	1138.3	1156
Jiamusi	8178	1289.9	1271
Mudanjiang	38755	1055.4	1315
Yichun	2860	1262.5	1387
Qitaihe	290	1283	1460
Jixi	2740	1223	1509
Xilinguole league	11208	1026.1	1513
Hegang	467	1354	1521
Shuangyashan	2344	1358.1	1544
Heihe	5241	1510.1	1582
Daxinganling	230	1699.1	1587
Hulunbeir	32944	1474.6	1663

Northeast China, and the highest dry port grade that each city can build is obtained. Using SPSS for dimension reduction factor analysis, the result of the KMO test is 0.868, and the probability of Bartlett's sphericity test statistical value is less than 0.001 ($P \leq 0.001$), which shows that the selected variables have high correlation and are suitable for factor analysis. In accordance with the principle of eigenvalues greater than 1, two common factors are extracted by principal component analysis, and the cumulative contribution rate of the principal component is 87.457%. Table 8 shows the scores and grade limits of the construction capacity of the dry ports in each city.

5.4. Model Solution Results. In this study, we use the AMPL software, select the optimizer Gurobi 9.0.2 to solve the model, and obtain the cost of different transportation modes and the construction scale level, quantity, and hinterland range of the dry port. Moreover, this study analyzes the model results from two aspects: transportation cost change and the dry port construction scheme.

5.4.1. Changes in Transportation Costs. From the perspective of total cost, the total cost of direct transportation of containers by road is the highest, at about 1,202.23 million USD, and the total cost of the RD-RL mode is the lowest, at approximately 941.68 million USD (Table 9). Compared with the RD mode, the total costs of the RD–RD-RL mode (container cannot be split), RD-RL–RD mode, and RD-RL mode are lower by 17.85%, 18.92%, and 21.67%, respectively. The results reveal that the container multimodal transport mode of dry port transshipment can effectively reduce the total transportation cost. The RD-RL mode is the best transport mode with the optimal total transport cost. This indicates that whether the railway can reach the seaport directly has a significant impact on reducing the cost of container transportation.

Regarding transportation economic cost, the proportion of economic cost under different transportation modes in the total transportation cost can reach more than 88%, and the economic cost of the RD mode is the highest. Compared with the RD mode, the economic costs of the RD-RL-RD (container cannot be split), RD-RL-RD, and RD-RL modes are lower by 18.03%, 19.82%, and 22.98%, respectively.

Table 5: Railway transport related charging standards.

Pay service	Charge standard (20-foot box)	Charging basis
Base price of railway delivery (P^1)	65.14 USD/TEU	China railway Corporation: Notice on the adjustment of railway container freight rate (2018)
Base price of railway operation (P^2)	0.47 USD (TEU·km)	
Container usage fee (P^3)	If it is less than 500 km, it's 19.25 USD/TEU; if it is less than 2000 km, it in 1.92 USD per 100 km	Notice on adjusting container usage fee and freight tarpaulin usage fee (tyy (2008) No. 144)
Container cleaning fee (P^4)	0.74 USD/TEU	Notice on adjusting the rates of some passenger and freight miscellaneous charges and announcing the rates of railway coal dust suppression transportation and items (T. Y. (2009) No. 224)
Container handling charge (P^5)	28.87 USD/TEU	Article 16 of the notice of the Ministry of railways on revising and re promulgating the charging method for railway cargo handling operations
Unloading fee of container in yard (P^6)	26.65 USD/TEU	

Table 6: Quadratic function regression results of road transportation economic cost.

Dependent variable: Y

Method: least squares

Date: 01/26/21 time: 16 : 19

Sample: 1 83

Included observations: 83

$Y = C(1) * X2 + C(2) * X + C(3)$

Coefficient	Std. Error	t-statistic	Prob.	
$C(1)$	0.000733	0.000132	5.542924	0
$C(2)$	5.753184	0.430664	13.35888	0
$C(3)$	914.8968	175.7349	5.206118	0
R-squared	0.960057	Mean dependent var		4063.855
Adjusted R-squared	0.959059	S.D. dependent var		4921.867
S.E. of regression	995.8865	Akaike info criterion		16.68062
Sum squared resid	79343192	Schwarz criterion		16.76805
Log likelihood	-689.2457	Hannan-Quinn criter.		16.71574
F-statistic	961.4377	Durbin-Watson stat		1.765773
Prob (F-statistic)	0			

In terms of time cost, the impact of the time cost of the different transportation modes on the total transportation cost is almost negligible (Table 10). The time cost of the RD mode is the lowest. Compared with the RD mode, the time cost of the RD-RL-RD (container cannot be split), RD-RL-RD, and RD-RL modes are higher by 4.63%, 7.12%, and 8.02%, respectively. This indicates that the direct road transportation mode has greater competitive advantage in terms of transportation time.

With regard to carbon emission cost, the carbon emission cost of the RD mode accounts for 10.82% of the total cost, while that of the multimodal transportation mode accounts for less than 5%. Compared with the RD mode, the carbon emission costs of the RD-RL-RD (container cannot be split), RD-RL-RD, and RD-RL modes are lower by 63.06%, 76.34%, and 78.37%, respectively. This indicates that multimodal transport can effectively reduce carbon emissions and the associated costs.

5.4.2. Dry Port Construction. Tables 8 and 9 show the scale grade, construction quantity, location layout, and hinterland distribution of the dry port under the multimodal transport mode of dry port transshipment. Under the RD-RL-RD mode, seven dry ports need to be built in Northeast China, including two fourth-level dry ports: Shenyang and Changchun, with a utilization rate of 100%; one third-level dry port: Harbin, with a utilization rate of 100%; one second-level dry port: Yanbian Korean Autonomous Prefecture, with a utilization rate of 54.17%; and three first-level dry ports: Chaoyang, Liaoyuan, and Hulunbuir, with an average utilization rate of 83.30%, all of which carry 79.88% of the container transshipment volume. Under the RD-RL mode with the optimal total transportation cost, the utilization rate of the four-level dry ports is 100% in Shenyang and Changchun, 100% in Harbin, 73.33% in Siping and Yanbian Korean Autonomous Prefecture, and 74.95% in Chaoyang and Hulunbuir. In this case, the dry ports carry 82.76% of the container transshipment volume. If the railway can reach the seaport directly, the proportion of the containers transshipped through the dry port in Shenyang will increase by 13.39%. The newly built second-level dry port Siping will replace Liaoyuan, the first-level dry port transshipping all the local containers and sharing part of the containers

Table 7: Classification of dry port scale.

Level of dry port	Construction cost (million USD)	Operating cost (million USD)	Designed carrying capacity (10,000 TEU)	Storage waiting time (day)	Discount coefficient of the scale effect
1	19.25	1.11	5	2	1
2	38.49	2.22	10	1.8	0.9
3	76.98	4.44	20	1.5	0.8
4	153.97	8.88	40	1	0.7

Table 8: Score and grade limit of dry port construction capacity of each city.

City	Score	Maximum level of dry port
Shenyang	2.51	4
Harbin	2.11	4
Changchun	1.84	4
Jilin	0.37	3
Daqing	0.36	3
Anshan	0.17	3
Mudanjiang	0.11	3
Chifeng	0.02	3
Qiqihar	−0.02	2
Benxi	−0.04	2
Hulunbuir	−0.07	2
Songyuan	−0.11	2
Tonghua	−0.11	2
Yanbian	−0.14	2
Jiamusi	−0.14	2
Tongliao	−0.16	2
Chaoyang	−0.19	2
Siping	−0.22	2
Suihua	−0.22	2
Xilin Gol League	−0.22	2
Fuxin	−0.28	1
Baicheng	−0.29	1
Liaoyang	−0.33	1
Liaoyuan	−0.33	1
Jixi	−0.35	1
Fushun	−0.36	1
Baishan	−0.38	1
Tieling	−0.38	1
Heihe	−0.41	1
Xing'an League	−0.42	1
Shuangyashan	−0.43	1
Hegang	−0.44	1
Daxinganling	−0.47	1
Qitaihe	−0.47	1
Yichun	−0.50	1

transshipped through Shenyang dry port in RD-RL-RD mode. This shows that the "last kilometer" problem has a great impact on the distribution of cargo demand, and to a certain extent, it also affects the location and layout of the dry port.

The locations of the dry ports and the spatial layout of the hinterland in the RD-RL-RD and RD-RL modes are shown in Figures 2 and 3. In the two transportation modes, the hinterland scope of the dry port is basically the same, and the location layout of the dry port is relatively decentralized and balanced, covering more than 85% of the inland cities in Northeast China. The hinterland cities of the dry ports are all located away from the harbor, except for the city itself. This indicates that there is no circuitous transportation away from the seaport when containers are transferred through the dry port. The competition between some dry ports is fierce, in the RD-RL-RD mode. Mudanjiang city is the competitive hinterland of Yanbian dry port and Harbin dry port; Changchun City is the competitive hinterland of Changchun dry port and Liaoyuan dry port; and Siping City is the competitive hinterland of Shenyang dry port and Liaoyuan dry port. In the RD-RL-RD mode, Changchun City is the competitive hinterland of Changchun dry port and Siping dry port Tables 10–12.

Table 9: Cost changes under the different transportation modes (unit: USD).

Cost type	RD mode	RD-RL-RD mode (container cannot be split)	RD-RL-RD mode	RD-RL mode
C	1,202,232,613.78	987,667,488.93	974,822,985.02	941,684,421.50
C^{rd}	1,070,569,856.88 (89.048%)	317,694,704.07 (32.166%)	145,297,084.94 (14.905%)	125,104,018.01 (13.285%)
C^{rdt}	1,581,666.09 (0.132%)	434,051.05 (0.044%)	198,740.18 (0.020%)	171,304.69 (0.018%)
C^{rdc}	130,081,090.81 (10.82%)	34,035,122.82 (3.446%)	14,337,233.56 (1.471%)	12,040,588.60 (1.279%)
$\sum_{i\in I}\sum_{j\in J}C_{ij}^{rd}$	—	194,490,393.74 (19.692%)	259,444,699.67 (26.615%)	263,679,554.78 (28.001%)
$\sum_{i\in I}\sum_{j\in J}C_{ij}^{rdt}$	—	122,188.21 (0.012%)	140,500.30 (0.014%)	135,410.69 (0.014%)
$\sum_{i\in I}\sum_{j\in J}C_{ij}^{rdt}$	—	10,451,861.70 (1.058%)	12,083,217.50 (1.240%)	11,669,116.14 (1.239%)
$\sum_{i\in I}\sum_{j\in J}\sum_{m\in M}C_{ij}^{rl}$	—	340,994,766.19 (34.525%)	418,724,637.15 (42.954%)	435,727,797.95 (46.271%)
$\sum_{i\in I}\sum_{j\in J}\sum_{m\in M}C_{ij}^{rlt}$	—	1,091,091.95 (0.111%)	1,344,020.64 (0.138%)	1,401,783.50 (0.149%)
$\sum_{i\in I}\sum_{j\in J}\sum_{m\in M}C_{ij}^{rlc}$	—	3,484,266.52 (0.353%)	4,250,156.61 (0.436%)	4,425,452.51 (0.47%)
$\sum_{i\in I}\sum_{j\in J}C_{kk}^{rd}$	—	24,326,913.11 (2.463%)	34,912,549.78 (3.581%)	—
$\sum_{i\in I}\sum_{j\in J}C_{kk}^{rdt}$	—	7,636.87 (0.001%)	10,959.99 (0.001%)	—
$\sum_{i\in I}\sum_{j\in J}C_{kk}^{rdc}$	—	75,681.03 (0.008%)	108,612.94 (0.011%)	—
$C_c + C_o$	—	60,458,811.66 (6.121%)	83,970,571.75 (8.614%)	87,329,394.62 (9.274%)

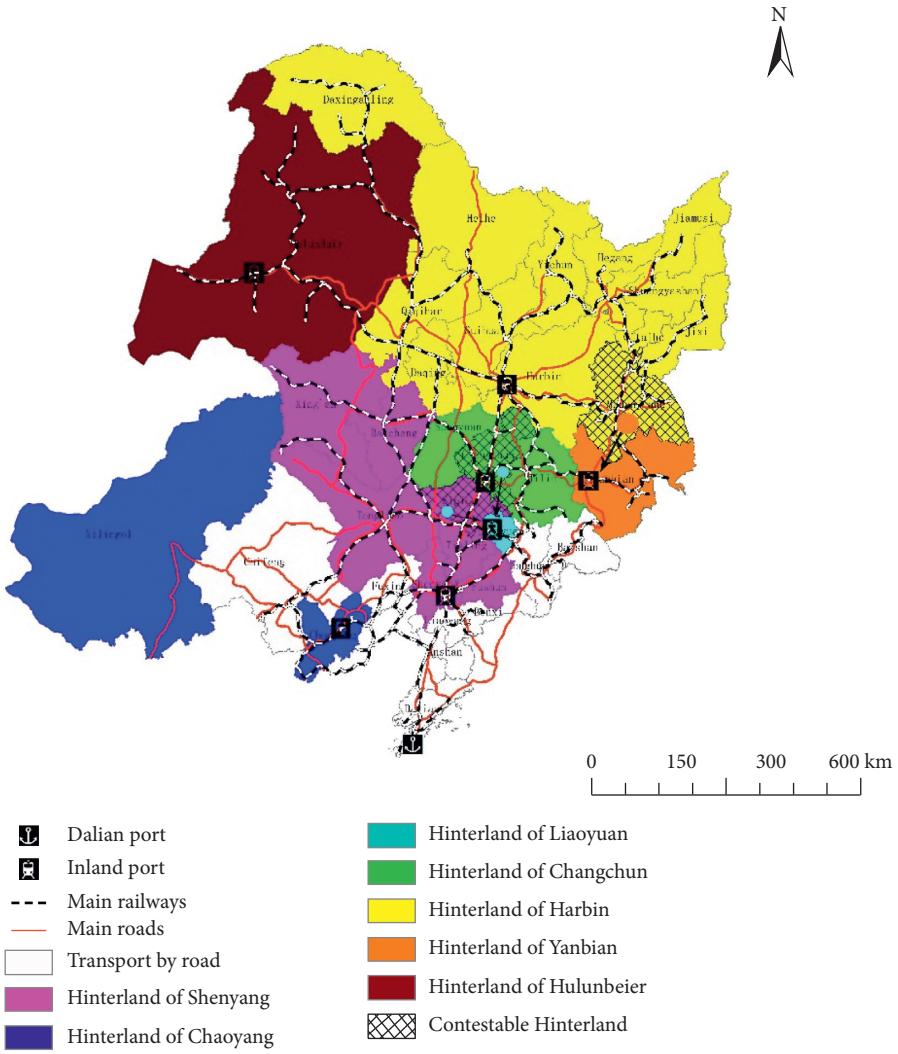


Figure 2: Hinterland area of the dry port under the road-rail-road mode.

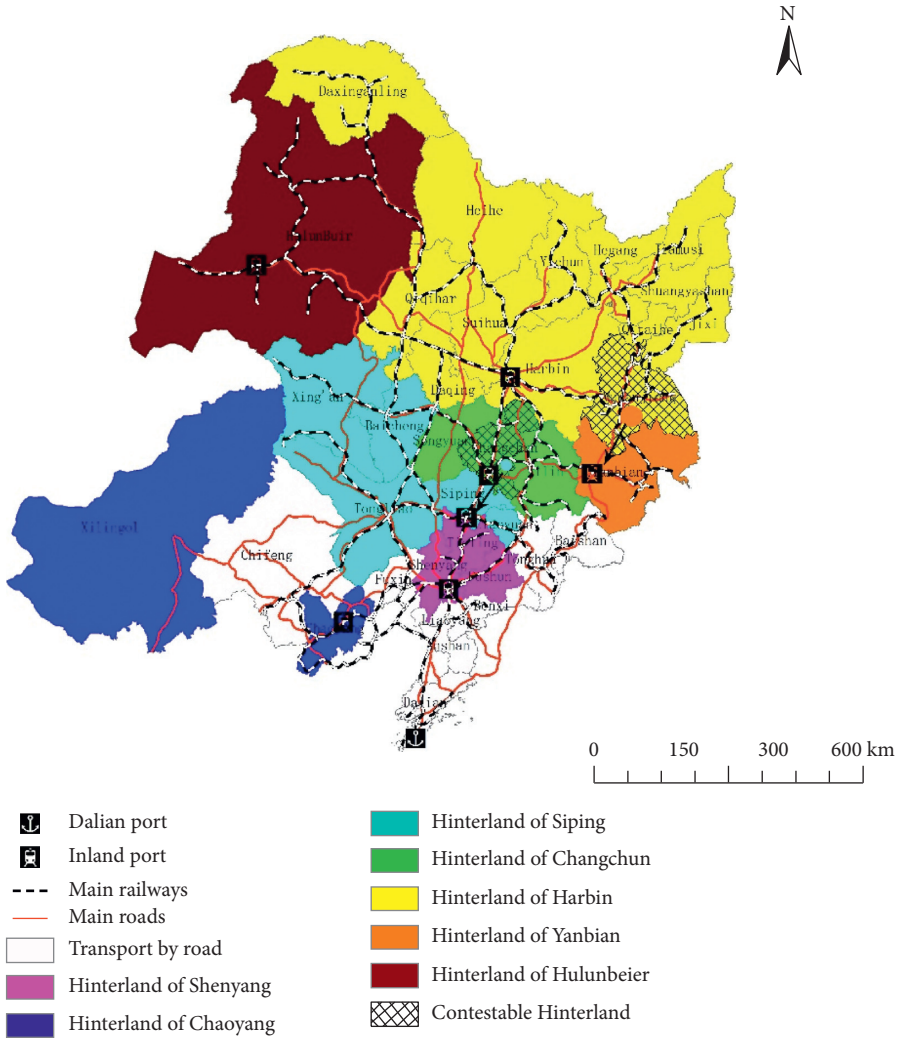


Figure 3: Hinterland area of the dry port under the road-rail mode.

Table 10: Cost of the different transportation modes and their proportion in total cost (unit: USD).

	RD mode	RD-RL-RD mode (container cannot be split)	RD-RL-RD mode	RD-RL mode
Economic cost	1,070,569,856.88 (89.048%)	877,506,777.12 (88.846%)	858,378,971.54 (88.055%)	824,511,370.74 (84.581%)
Time cost	1,581,666.09 (0.132%)	1,654,968.08 (0.168%)	1,694,221.12 (0.174%)	1,708,498.89 (0.175%)
Carbon emission cost	130,081,090.81 (10.82%)	48,046,932.06 (4.865%)	30,779,220.61 (3.157%)	28,135,157.24 (2.886%)
Construction and operation cost of dry port	—	60,458,811.66 (6.121%)	83,970,571.75 (8.614%)	87,329,394.62 (8.958%)

Table 11: Construction of the dry port under the RD-RL-RD mode.

Serial number	City name	Level of dry port	Transshipment volume (TEU)	Number of hinterland cities	Hinterland cities and transshipment volume (TEU)
1	Shenyang	4	400,000	7	Shenyang (317,186), Fushun (31,920), Tieling (8,408), Siping (8,982), Tongliao (17,664), Baicheng (15,547), and Xing'an (293)
2	Chaoyang	1	42,008	2	Chaoyang (30,800) and Xilin Gol (11,208)
3	Liaoyuan	1	50,000	3	Liaoyuan (7,254), Siping (781), and Changchun (41,965)
4	Changchun	4	400,000	3	Changchun (353,986), Jilin (36,360), and Songyuan (9,654)
5	Harbin	3	200,000	13	Harbin (45,141), Suihua (12,569), Daqing (81,033), Qiqihar (5,300), Jiamusi (8,178), Mudanjiang (33,607), Yichun (2,860), Qitaihe (290), Jixi (2,740), Hegang (467), Shuangyashan (2,344), Heihe (5241), and Daxinganling (230)
6	Yanbian	2	54,167	2	Yanbian (49,019) and Mudanjiang (5,148)
7	Hulun Buir	1	32,944	1	Hulun Buir (32,944)

Table 12: Construction of the dry port under the road-rail mode.

Serial number	City name	Level of dry port	Transshipment volume (TEU)	Number of hinterland cities	Hinterland cities and transshipment volume (TEU)
1	Shenyang	4	400,000	3	Shenyang (359,672), Fushun (31,920), and Tieling (8,408)
2	Chaoyang	1	42,008	2	Chaoyang (30,800) and Xilin Gol (11,208)
3	Siping	2	92,486	6	Liaoyuan (7,254), Siping (9,763), Tongliao (17,664), Changchun (41,965), Baicheng (15,547), and Xing'an (293)
4	Changchun	4	400,000	3	Changchun (353,986), Jilin (36,360), and Songyuan (9,654)
5	Harbin	3	200,000	13	Harbin (45,141), Suihua (12,569), Daqing (81,033), Qiqihar (5,300), Jiamusi (8,178), Mudanjiang (33,607), Yichun (2,860), Qitaihe (290), Jixi (2,740), Hegang (467), Shuangyashan (2,344), Heihe (5,241), and Daxinganling (230)
6	Yanbian	2	54,167	2	Yanbian (49,019) and Mudanjiang (5,148)
7	Hulun Buir	1	32,944	1	Hulun Buir (32,944)

6. Conclusion and Discussion

This study uses a technical economic method to calculate the construction cost of a dry port by year and takes the operation and time costs of the dry port into the comprehensive cost. Considering the scale effect of railway transportation, the capacity level of the dry port and the conditions of urban construction of the dry port, this study constructs a multicapacity-level model for the location of the dry port. From an empirical analysis of the situation in Northeast China using the AMPL software solution, the following conclusions are made.

(1) In the context of rapid economic development and stabilization of transport demand, dry port services can effectively reduce transport costs and carbon emission costs, lower the total cost, realize economies of scale in transport, and promote the transformation of the RD mode to a cleaner mode of transport.

(2) The RD-RL mode is the optimal method of container transportation. In this case, seven dry ports need to be built in Northeast China—two fourth-grade dry ports: in Shenyang and Changchun, with a utilization rate of 100%; one third-grade dry port: in Harbin, with a utilization rate of 100%; two second-grade dry ports: in Siping and Yanbian, with an average utilization rate of 73.33%; and two first-grade dry ports: in Chaoyang and Hulunbuir, with an average utilization rate of 74.95%, carrying 82.76% of the container transshipment volume. Compared with the RD mode, the total, economic, and carbon emission costs of the RD-RL mode are lower by 21.67%, 22.98%, and 78.37%, respectively, and the time cost is higher by 8.02%.

(3) Considering that the container cargo can be split, it has a great impact on the distribution of freight flow and transportation cost, and if only the container in the shipper city can be split at will in the transportation process, the total transportation costs can be optimized.

(4) In the RD-RL-RD mode, the road transportation cost between the shipper and the dry port accounts for 27.869% of the total transportation cost, and the road transportation cost between the seaport railway station and the seaport accounts for 3.593% of the total cost. The road transportation cost accounts for 31.462% of the total transportation cost. In the RD-RL mode, the road transport cost between the shipper and the dry ports accounts for 29.254% of the total transport cost. This indicates that the road

transportation cost between the shipper and the dry port has a significant impact on the total transportation cost, freight demand distribution, and dry port selection. Compared with the RD-RL-RD mode, the total annual transport cost of the RD-RL mode is lower by approximately 3.340%, CO_2 emissions are lower by approximately 8.590%, and the dry port selection scheme does not change significantly. This indicates that although the direct rail connection to the seaport does not have a significant impact on the dry port selection scheme, it helps in reducing the total transportation cost, especially the negative impact on the environment.

The market-oriented reform of China's railway operations has been intensifying, and the development of multimodal transport has been accelerating. This study considers the scale effect (discount coefficient) of railway transportation to make the study of dry port locations more consistent with the actual development. In addition, the location selection process in this study considers both the capacity level of the dry ports and the conditions for cities to build dry ports, which is an improvement of the two-stage dry port location method. It ensures that the number, level, and layout of dry ports are in line with the actual situation.

Future research may explore the following four aspects: First, the problem of dry port location is a complex system project involving many influential factors. More factors (such as weight coefficient of different costs) can be considered while setting optimization objectives, and the limitation of railway capacity can be considered while setting constraints. Second, this study has solved the joint optimization problem of land port planning and container transportation for a specific port, and the land port location problem may need to be extended to multiple ports. Third, this study focuses on regional-level network and intends to solve the optimization problem of the nationwide network. Fourth, this study focuses on the macro spatial layout of land port planning, and future research may be extended to the microlayout of the land port in the urban planning of the land port city.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was funded by the Key Project of Natural Science Foundation of China, grant number 42030409, the Youth Program of National Natural Science Foundation of China, grant number 41701134, the National Natural Science Foundation of China, grant number 41671117, and the Postdoctoral Science Foundation of China, grant number 2019M651098.

References

[1] R. Wang, Q. Sun, W. Hu, Y. Li, D. Ma, and P. Wang, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, p. 1, 2021.

[2] Q. Sun, R. Han, H. Zhang, J. Zhou, and J. M. Guerrero, "A multiagent-based consensus algorithm for distributed coordinated control of distributed generators in the energy internet," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 3006–3019, 2015.

[3] X. Hu, H. Zhang, D. Ma, and R. Wang, "A TnGAN-based leak detection method for pipeline network considering incomplete sensor data," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.

[4] V. Roso, J. Woxenius, and K. Lumsden, "The dry port concept: connecting container seaports with the hinterland," *Journal of Transport Geography*, vol. 17, no. 5, pp. 338–345, 2009.

[5] M. Gonzalez-Aregall and R. Bergqvist, "The role of dry ports in solving seaport disruptions: a Swedish case study," *Journal of Transport Geography*, vol. 80, Article ID 102499, 2019.

[6] V. Roso, "Evaluation of the dry port concept from an environmental perspective: a note," *Transportation Research Part D: Transport and Environment*, vol. 12, no. 7, pp. 523–527, 2007.

[7] L. Lättilä, V. Henttu, and O.-P. Hilmola, "Hinterland operations of sea ports do matter: dry port usage effects on transportation costs and CO2 emissions," *Transportation Research Part E: Logistics and Transportation Review*, vol. 55, pp. 23–42, 2013.

[8] A. Carboni and F. Orsini, "Dry ports and related environmental benefits: a case study in Italy," *Case Studies on Transport Policy*, vol. 8, no. 2, pp. 416–428, 2020.

[9] X. Qiu and C.-Y. Lee, "Quantity discount pricing for rail transport in a dry port system," *Transportation Research Part E: Logistics and Transportation Review*, vol. 122, pp. 563–580, 2019.

[10] Analysis on the market situation and competition pattern of China's port industry in 2020, 2020. https://www.sohu.com/a/438868608_473133.

[11] F. Liu and H. Xu, "Heterogeneity of green TFP in China's logistics industry under environmental constraints," *Complexity*, vol. 2020, Article ID 8842560, 2020.

[12] H. Wang, J. Y. Han, M. Su, S. L. Wan, and Z. C. Zhang, "The relationship between freight transport and economic development: a case study of China," *Research in Transportation Economics*, vol. 85, Article ID 100885, 2020.

[13] T. Van Nguyen, J. Zhang, L. Zhou, M. Meng, and Y. He, "A data-driven optimization of large-scale dry port location using the hybrid approach of data mining and complex network theory," *Transportation Research Part E: Logistics and Transportation Review*, vol. 134, Article ID 101816, 2020.

[14] Z. Chang, D. Yang, Y. Wan, and T. Han, "Analysis on the features of Chinese dry ports: ownership, customs service, rail service and regional competition," *Transport Policy*, vol. 82, pp. 107–116, 2019.

[15] S.-J. Jeong, C.-G. Lee, and J. H. Bookbinder, "The European freight railway system as a hub-and-spoke network," *Transportation Research Part A: Policy and Practice*, vol. 41, no. 6, pp. 523–536, 2007.

[16] W. F. Wang and W. Y. Yun, "Scheduling for inland container truck and train transportation," *International Journal of Production Economics*, vol. 143, no. 2, pp. 349–356, 2013.
[17] E. Demir, W. Burgholzer, M. Hrušovský, E. Arıkan, W. Jammernegg, and T. V. Woensel, "A green intermodal service network design problem with travel time uncertainty," *Transportation Research Part B: Methodological*, vol. 93, pp. 789–807, 2016.
[18] R. Wang, K. Yang, L. Yang, and Z. Gao, "Modeling and optimization of a road-rail intermodal transport system under uncertain information," *Engineering Applications of Artificial Intelligence*, vol. 72, pp. 423–436, 2018.
[19] Y. Zhao, Z. Yang, and H. Haralambides, "Optimizing the transport of export containers along China's coronary artery: the Yangtze River," *Journal of Transport Geography*, vol. 77, pp. 11–25, 2019.
[20] X. Feng, Y. Zhang, Y. Li, and W. Wang, "A location-allocation model for seaport-dry port system optimization," *Discrete Dynamics in Nature and Society*, vol. 2013, Article ID 309585, 9 pages, 2013.
[21] Z. Chang, Z. Qi, and J. Lu, "Study on the location of inland port with multiple capacity levels selection," *Transportation System Engineering and Information*, vol. 17, no. 3, pp. 184–191, 2017.
[22] Y.-C. Tsao and V. T. Linh, "Seaport-dry port network design considering multimodal transport and carbon emissions," *Journal of Cleaner Production*, vol. 199, pp. 481–492, 2018.
[23] X. Xu, Q. Zhang, W. Wang, Y. Peng, X. Song, and Y. Jiang, "Modelling port competition for intermodal network design with environmental concerns," *Journal of Cleaner Production*, vol. 202, pp. 720–735, 2018.
[24] Q. Zhang, W. Wang, Y. Peng, J. Zhang, and Z. Guo, "A game-theoretical model of port competition on intermodal network and pricing strategy," *Transportation Research Part E: Logistics and Transportation Review*, vol. 114, pp. 19–39, 2018.
[25] X. Jiang, H. Fan, M. Luo, and Z. Xu, "Strategic port competition in multimodal network development considering shippers' choice," *Transport Policy*, vol. 90, pp. 68–89, 2020.
[26] H. Wei and M. Dong, "Import-export freight organization and optimization in the dry-port-based cross-border logistics network under the belt and road initiative," *Computers & Industrial Engineering*, vol. 130, pp. 472–484, 2019.
[27] Y.-C. Tsao and V.-V. Thanh, "A multi-objective mixed robust possibilistic flexible programming approach for sustainable seaport-dry port network design under an uncertain environment," *Transportation Research Part E: Logistics and Transportation Review*, vol. 124, pp. 13–39, 2019.
[28] L. H. Kaack, P. Vaishnav, M. G. Morgan, I. L. Azevedo, and S. Rai, "Decarbonizing intraregional freight systems with a focus on modal shift," *Environmental Research Letters*, vol. 13, no. 8, Article ID 083001, 2018.
[29] M. E. O'Kelly, "Hub facility location with fixed costs," *Papers in Regional Science*, vol. 71, no. 3, pp. 293–306, 1992.
[30] M. Janic, "Modelling the full costs of an intermodal and road freight transport network," *Transportation Research Part D: Transport and Environment*, vol. 12, no. 1, pp. 33–44, 2007.
[31] M. Zhang, B. Wiegmans, and L. Tavasszy, "Optimization of multimodal networks including environmental costs: a model and findings for transport policy," *Computers in Industry*, vol. 64, no. 2, pp. 136–145, 2013.
[32] T.-S. Chang, "Best routes selection in international intermodal networks," *Computers & Operations Research*, vol. 35, no. 9, pp. 2877–2891, 2008.
[33] E. Pekin, C. Macharis, D. Meers, and P. Rietveld, "Location analysis model for Belgian intermodal terminals: importance of the value of time in the intermodal transport chain," *Computers in Industry*, vol. 64, no. 2, pp. 113–120, 2013.
[34] Y. Rahimi, R. Tavakkoli-Moghaddam, M. Mohammadi, and M. Sadeghi, "Multi-objective hub network design under uncertainty considering congestion: an M/M/c/K queue system," *Applied Mathematical Modelling*, vol. 40, no. 5-6, pp. 4179–4198, 2016.
[35] N. Mithulananthan and N. Acharya, "A proposal for investment recovery of FACTS devices in deregulated electricity markets," *Electric Power Systems Research*, vol. 77, no. 5-6, pp. 695–703, 2007.
[36] JCTRANS. 2020, http://land.jctrans.com/.
[37] H. Li and L. Su, "Multimodal transport path optimization model and algorithm considering carbon emission multitask," *The Journal of Supercomputing*, vol. 76, no. 12, pp. 9355–9373, 2020.
[38] Z. Chang, T. Notteboom, and J. Lu, "A two-phase model for dry port location with an application to the port of Dalian in China," *Transportation Planning and Technology*, vol. 38, no. 4, pp. 442–464, 2015.

Hindawi
Complexity
Volume 2021, Article ID 5549678, 10 pages
https://doi.org/10.1155/2021/5549678

Research Article

Neural Network-Based Intelligent Computing Algorithms for Discrete-Time Optimal Control with the Application to a Cyberphysical Power System

Feng Jiang, Kai Zhang, Jinjing Hu, and Shunjiang Wang

State Grid Liaoning Electric Power Company Limited, Shenyang 110006, China

Correspondence should be addressed to Shunjiang Wang; wangshunjiang@163.com

Received 26 February 2021; Revised 14 March 2021; Accepted 7 April 2021; Published 17 April 2021

Academic Editor: Qiuye Sun

Adaptive dynamic programming (ADP), which belongs to the field of computational intelligence, is a powerful tool to address optimal control problems. To overcome the bottleneck of solving Hamilton–Jacobi–Bellman equations, several state-of-the-art ADP approaches are reviewed in this paper. First, two model-based offline iterative ADP methods including policy iteration (PI) and value iteration (VI) are given, and their respective advantages and shortcomings are discussed in detail. Second, the multistep heuristic dynamic programming (HDP) method is introduced, which avoids the requirement of initial admissible control and achieves fast convergence. This method successfully utilizes the advantages of PI and VI and overcomes their drawbacks at the same time. Finally, the discrete-time optimal control strategy is tested on a power system.

1. Introduction

Adaptive dynamic programming (ADP) [1–4], which integrates the advantages of reinforcement learning (RL) [5–8] and adaptive control, has become a powerful tool in solving optimal control problems. With decades of development, ADP has also provided many approaches to solve other control problems, such as robust control [9, 10], optimal control with input constraints [11, 12], optimal tracking control [13, 14], zero-sum games [15], and non-zero-sum games [16]. Furthermore, ADP methods have been widely applied to the real-world systems, such as water-gas shift reaction [17], battery management [18], microgrid systems [19, 20], and Quanser helicopter [21]. These aforementioned papers were all inspired and developed by the basic works of ADP-based optimal control; i.e., optimal control is the core research topic of ADP.

The bottleneck of solving the nonlinear optimal control problems is to obtain the solutions of Hamilton–Jacobi–Bellman (HJB) equations. However, these equations are generally difficult or even impossible to be solved analytically. To overcome this difficulty, ADP has given several important iterative learning frameworks, such as policy iteration (PI) [2, 22, 23] and value iteration (VI) [24–26]. PI algorithm starts from an initial admissible control policy and then proceeds the policy evaluation step and the policy improvement step successively till convergence. The main advantage of PI is that it ensures all the iterative control policies are admissible and achieves fast convergence. The drawback of PI is also obvious. The requirement of initial admissible control is a strict condition in practice, which seriously limits its applications. Different from PI, VI can start from an arbitrary-positive semidefinite value function, which is an easy-to-realize initial condition. Although the easier initial condition makes VI more practical, it also leads to a longer iteration learning process; that is, VI achieves convergence much slower than PI. Thus, it is desired to develop a new method, which avoids the requirement of initial admissible control and gets convergence faster than the VI algorithm. To realize these purposes, the multistep heuristic dynamic programming (HDP) approach [27] is presented to integrate the merits of PI and VI algorithms and overcome their drawbacks.

This paper reviews the state-of-the-art ADP algorithms for the optimal control of discrete-time (DT) systems. The rest of this paper is arranged as follows. In Section 2, the problem formulation is derived. Three iterative model-based offline learning algorithms along with comprehensive comparisons are presented in Sections 3 and 4. The proposed DT optimal control strategy is tested on a power system in Section 5. Finally, a brief conclusion is drawn in Section 6.

2. Problem Formulation

In this paper, we consider the general nonlinear DT system:

$$x(k+1) = f(x(k)) + g(x(k))u(k), \tag{1}$$

where $x(k) \in \mathbb{R}^n$ represents the system state, $u(k) \in \mathbb{R}^m$ denotes the control input, and $f(x) \in \mathbb{R}^n$ and $g(x) \in \mathbb{R}^{n\times m}$ are the system functions.

The purpose of the optimal control issue is to find out a state feedback control policy $u(x(k))$, which can not only stabilize system (1) but also minimize the following performance index function:

$$J(x(0), u) = \sum_{p=0}^{\infty} r(x(p), u(p)), \tag{2}$$

where $r(x,u) = x^T Q x + u^T R u$. The matrices Q and R determine the performance of system states and control inputs, respectively. Given the admissible control policy $u(x(k))$, the value function can be described by

$$\begin{aligned} V(x(k)) &= \sum_{p=k}^{\infty} r(x(p), u(p)) = r(x(k), u(k)) \\ &\quad + \sum_{p=k+1}^{\infty} r(x(p), u(p)) \\ &= r(x(k), u(k)) + V(x(k+1)). \end{aligned} \tag{3}$$

According to the definition of optimal control, the optimal value function can be defined by

$$\begin{aligned} V^*(x(k)) &= \min_{u(x)} \sum_{p=k}^{\infty} r(x(p), u(p)) \\ &= \min_{u(x)} \{ r(x(k), u(k)) + V^*(x((k+1))) \}. \end{aligned} \tag{4}$$

By using the stationarity condition [28], the optimal control policy can be derived as

$$\begin{aligned} u^*(x(k)) &= \arg\min_{u(x)} \{ r(x(k), u(k)) + V^*(x(k+1)) \} \\ &= \arg\min_{u(x)} \{ r(x(k), u(k)) + V^*(f(x(k)) \\ &\quad + g(x(k))u(k)) \} \\ &= -\frac{1}{2} R^{-1} g^T(x(k)) \nabla V^*(x(k+1)), \end{aligned} \tag{5}$$

where $\nabla V^*(x(k+1)) = \partial V^*(x(k+1))/\partial x(k+1)$.

The key to obtaining the optimal control policy $u^*(x(k))$ is to solve the following DT HJB equation [27]:

$$\begin{aligned} V^*(x(k)) &= r(x(k), u^*(k)) + V^*(x(k+1)) \\ &= \frac{1}{4} \nabla V^{*T}(x(k+1)) g(x(k)) R^{-1} g^T(x(k)) \nabla V^* \\ &\quad (x(k+1)) + x^T(k) Q x(k) + V^*(x(k+1)). \end{aligned} \tag{6}$$

Remark 1. Figure 1 provides the relationship and difference between discrete-time and continuous-time optimal control. The real-world systems generally exist in the continuous-time forms. After mathematical modeling, they are formulated by the continuous-time system models. Through sampling and discretization, the continuous-time system models are converted into the discrete-time ones. Therefore, the associated performance indexes and HJB equations of discrete-time systems are in the discretization forms compared with the continuous-time systems. The key to solving the discrete-time optimal control issue is the discrete-time HJB equation, which is a nonlinear partial difference equation. The existing works regarding continuous-time systems are much more than the ones regarding discrete-time systems. In order to overcome this bottleneck, several ADP learning algorithms along with their neural network (NN) implementations will be introduced.

3. Model-Based PI Algorithm for the Optimal Control Problem of DT Systems

In this section, the model-based PI algorithm along with its NN implementation will be introduced in detail. The model-based PI algorithm [2, 23] is shown in Algorithm 1.

The actor-critic dual-network structure with the gradient-descent updating law is employed to implement Algorithm 1. First, construct the critic NN to approximate the iterative value function:

$$\widehat{V}^{l,q}(x(k)) = \widehat{W}_c^{l,qT} \phi_c(x(k)), \tag{7}$$

where $\widehat{W}_c^{l,q}$ and $\phi_c(x)$ denote the NN weights and NN activation functions of the critic network and q is the iteration index for the following gradient-descent method.

Define the error function for the critic NN:

$$\widetilde{V}^{l,q}(x(k)) = \widehat{V}^{l,q}(x(k)) - V^l(x(k)), \tag{8}$$

where $V^l(x(k)) = \sum_{h=0}^{H-1} r(x(k+h), u^l(k+h)) + V^l(x(k+H))$. If we select a large enough integer H, then, with the admissible control $u^l(k)$, one has $V^l(x(k+H)) \longrightarrow 0$ [2]; that is, $V^l(x(k))$ can be expressed as $V^l(x(k)) = \sum_{h=0}^{H-1} r(x(k+h), u^l(k+h))$.

In order to minimize the error performance $E_c^{l,q}(k) = (\widetilde{V}^{l,q}(x(k)))^2/2$, the gradient-descent-based updating law for the critic NN is given by

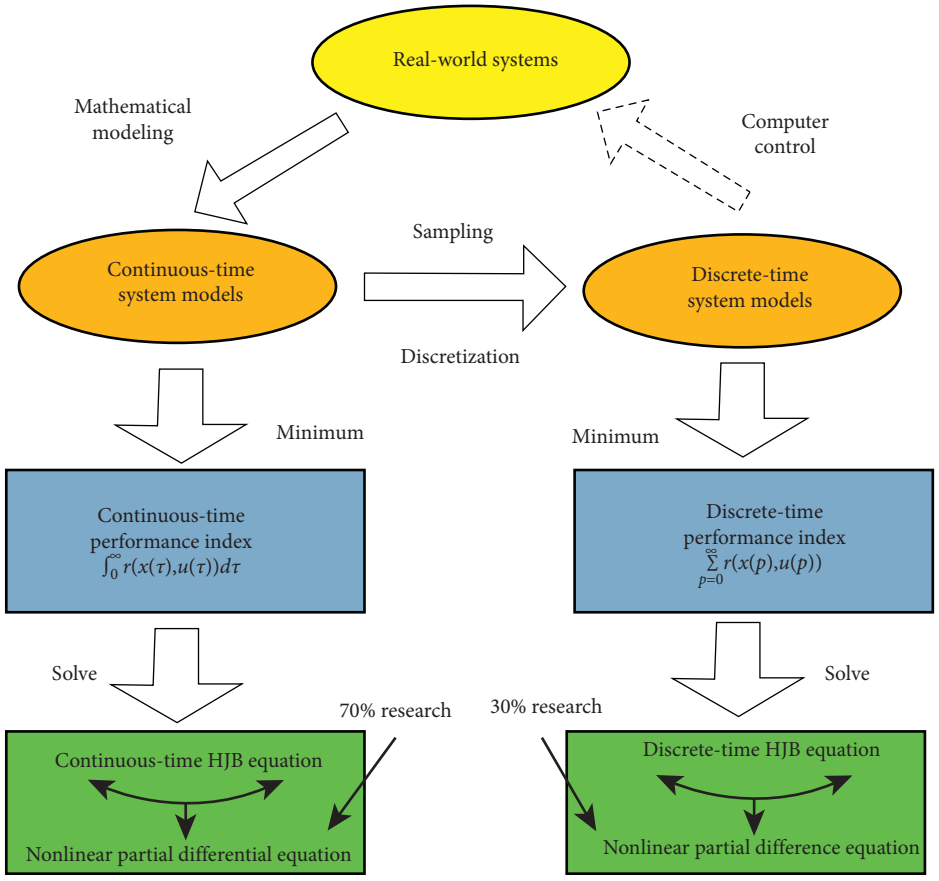


Figure 1: Relationship and difference between discrete-time and continuous-time optimal control.

Step 1: (Initialization)
Let the iteration index $l = 0$.
Select an initial admissible control policy $u^0(x)$.
Choose a small enough computation precision ϵ.
Step 2: (Policy Evaluation)
With $u^l(x)$, compute the iterative value function $V^l(x)$ by
$V^l(x(k)) = r(x(k), u^l(k)) + V^l(x(k+1))$.
Step 3: (Policy Improvement)
With $V^l(x)$, update the iterative control policy $u^{l+1}(x)$ by
$u^{l+1}(x(k)) = -(1/2)R^{-1}g^T(x(k))\nabla V^l(x(k+1))$.
Step 4: if $\|V^l - V^{l-1}\| \le \varepsilon$, stop and the optimal control policy $u^{l+1}(x)$ is acquired; Else, let $l = l + 1$ and go back to Step 2.

Algorithm 1: Model-based PI.

$$\widehat{W}_c^{l,q+1} = \widehat{W}_c^{l,q} - \beta_c \left[\frac{\partial E_c^{l,q}(k)}{\partial \widetilde{V}^{l,q}(x(k))} \frac{\partial \widetilde{V}^{l,q}(x(k))}{\partial \widehat{V}^{l,q}(x(k))} \frac{\partial \widehat{V}^{l,q}(x(k))}{\partial \widehat{W}_c^{l,q}} \right] = \widehat{W}_c^{l,q} - \beta_c \widetilde{V}^{l,q}(x(k)) \phi_c(x(k)), \tag{9}$$

where β_c is the learning rate of the critic NN.

Similar to the design of critic NN, the actor network, which is used to approximate the iterative control policy, is expressed as

$$\widehat{u}^{l,q}(x(k)) = \widehat{W}_a^{l,qT} \phi_a(x(k)). \tag{10}$$

The error function for the actor NN is defined as

$$\widetilde{u}^{l,q}(x(k)) = \widehat{u}^{l,q}(x(k)) - u^l(x(k)), \tag{11}$$

where $u^l(x(k))$ can be attained according to Algorithm 1.

To minimize the error performance $E_a^{l,q}(k) = \widetilde{u}^{l,qT}(x(k))\widetilde{u}^{l,q}(x(k))/2$, using the chain derivation rule, the updating law for the actor NN is designed by

$$\begin{aligned}\widehat{W}_a^{l,q+1} &= \widehat{W}_a^{l,q} - \beta_a\left[\frac{\partial E_a^{l,q}(k)}{\partial \widetilde{u}^{l,q}(x(k))}\frac{\partial \widetilde{u}^{l,q}(x(k))}{\partial \widehat{u}^{l,q}(x(k))}\frac{\partial \widehat{u}^{l,q}(x(k))}{\partial \widehat{W}_a^{l,q}}\right] \\ &= \widehat{W}_a^{l,q} - \beta_a\phi_a(x(k))\widetilde{u}^{l,qT}(x(k)),\end{aligned} \tag{12}$$

where β_a is the learning rate of the actor NN.

Remark 2. Figure 2 displays the NN implementation diagram of PI algorithm. First, NN weights of the actor network should be chosen to generate admissible control. Second, critic and actor networks are updated via the gradient-descent-based learning law to realize policy evaluation and improvement steps, respectively. After iteration, critic and actor networks achieve convergence, where the NN-based approximate optimal control can be obtained. Many stability proofs of the NN implementation procedure have been given in the existing works. Here, we introduce the following rigorous proof to demonstrate the optimality and convergence.

Theorem 1. *Let the target iterative value function and control policy be described by $V^l(x(k)) = W_c^{lT}\phi_c(x(k))$ and $u^l(x(k)) = W_a^{lT}\phi_a(x(k))$, respectively. Let the critic and actor NNs be updated via (9) and (12), respectively. If the learning rates β_c and β_a are selected to be appropriately small, then the NN weights $\widehat{W}_c^{l,q}$ and $\widehat{W}_a^{l,q}$ will asymptotically converge to the ideal values W_c^l and W_a^l, respectively.*

Proof. Let $\widetilde{W}_c^{l,q} = \widehat{W}_c^{l,q} - W_c^l$ and $\widetilde{W}_a^{l,q} = \widehat{W}_a^{l,q} - W_a^l$. According to (9) and (12), it can be acquired that

$$\begin{aligned}\widetilde{W}_c^{l,q+1} &= \widetilde{W}_c^{l,q} - \beta_c\widetilde{V}^{l,q}(x(k))\phi_c(x(k)), \\ \widetilde{W}_a^{l,q+1} &= \widetilde{W}_a^{l,q} - \beta_a\phi_a(x(k))\widetilde{u}^{l,qT}(x(k)),\end{aligned} \tag{13}$$

where $\widetilde{V}^{l,q}(x(k)) = \widetilde{W}_c^{l,qT}\phi_c(x(k))$ and $\widetilde{u}^{l,q}(x(k)) = \widetilde{W}_a^{l,qT}\phi_a(x(k))$.

Construct the following Lyapunov function candidate:

$$P\left(\widetilde{W}_c^{l,q}, \widetilde{W}_a^{l,q}\right) = tr\left\{\widetilde{W}_c^{l,qT}\widetilde{W}_c^{l,q} + \widetilde{W}_a^{l,qT}\widetilde{W}_a^{l,q}\right\}. \tag{14}$$

The difference of the Lyapunov function (14) can be derived as

$$\begin{aligned}\Delta P\left(\widetilde{W}_c^{l,q}, \widetilde{W}_a^{l,q}\right) &= tr\left\{\widetilde{W}_c^{l,q+1T}\widetilde{W}_c^{l,q+1} - \widetilde{W}_c^{l,qT}\widetilde{W}_c^{l,q}\right\} + tr\left\{\widetilde{W}_a^{l,q+1T}\widetilde{W}_a^{l,q+1} - \widetilde{W}_a^{l,qT}\widetilde{W}_a^{l,q}\right\} \\ &\le \beta_c\left\|\widetilde{V}^{l,q}(x(k))\right\|^2\left(\beta_c\left\|\phi_c(x(k))\right\|^2 - 2\right) \\ &\quad + \beta_a\left\|\widetilde{u}^{l,q}(x(k))\right\|^2\left(\beta_a\left\|\phi_a(x(k))\right\|^2 - 2\right).\end{aligned} \tag{15}$$

If the learning rates are selected to satisfy $\beta_c \le 2/\|\phi_c(x(k))\|^2$ and $\beta_a \le 2/\|\phi_a(x(k))\|^2$, then one has $\Delta P(\widetilde{W}_c^{l,q}, \widetilde{W}_a^{l,q}) \le 0$, which implies the NN weights $\widehat{W}_c^{l,q}$ and $\widehat{W}_a^{l,q}$ will asymptotically converge to the ideal values.

This completes the proof. □

4. Model-Based VI Algorithm and Multistep HDP Algorithm

With the help of the initial admissible control, the PI algorithm achieves fast convergence. However, the weakness of the PI algorithm is obvious. The PI algorithm requires the initial control policy to be admissible, which is a strict condition. How to find out an initial admissible control policy is still an open problem, which limits the real-world applications of the PI algorithm. To relax the strict condition, the model-based VI algorithm [24–26] is shown in Algorithm 2, where the initial condition becomes much easier.

Remark 3. Different from the PI algorithm, the VI algorithm does not require the initial admissible control, and one only needs to provide a specific initial value function, which makes the VI algorithm more practical in the real-world applications. However, without the help of the initial admissible control, the VI algorithm generally suffers from the low convergence speed. From the aforementioned content, it can be observed that the PI and VI algorithms have their own advantages and disadvantages. The PI algorithm can achieve fast convergence, while it requires an initial admissible control policy. The VI algorithm can start from an easy-to-realize initial condition, while it generally suffers from the low convergence speed. Thus, it is expected to design a new approach, which can make the trade-off between the PI algorithm and the VI algorithm.

That is, it is desired to develop an algorithm, which achieves convergence faster than the VI algorithm and does not require an initial admissible control policy. To realize this goal, the multistep HDP method [27] will be introduced in Algorithm 3.

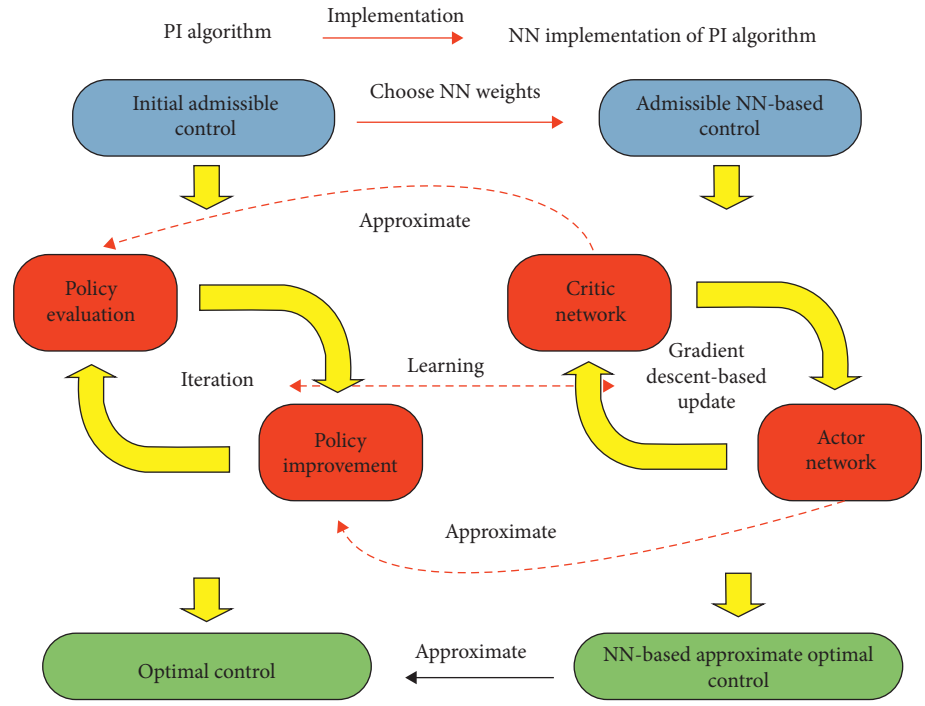


Figure 2: NN implementation diagram of PI algorithm.

Step 1: (Initialization)
Let the iteration index $l = 0$.
Select an initial value function $V^0(x)$.
Choose a small enough computation precision ϵ.
Step 2: (Policy Improvement)
With $V^l(x)$, compute the iterative control policy $u^l(x)$ by
$u^l(x(k)) = -(1/2)R^{-1}g^T(x(k))\nabla V^l(x(k+1))$.
Step 3: (Policy Evaluation)
With $u^l(x)$, calculate the iterative value function $V^{l+1}(x)$ by
$V^{l+1}(x(k)) = r(x(k), u^l(k)) + V^l(x(k+1))$.
Step 4: if $\|V^{l+1} - V^l\| \le \varepsilon$, stop and the optimal control policy $u^l(x)$ is acquired; Else, let $l = l + 1$ and go back to Step 2.

Algorithm 2: Model-based VI.

Construct the critic and actor NNs to approximate the iterative value function and control policy as follows:

$$\begin{aligned} \widehat{V}^l(x(k)) &= \phi_c^T(x(k))\widehat{W}_c^l, \\ \widehat{u}^l(x(k)) &= \phi_a^T(x(k))\widehat{W}_a^l, \end{aligned} \tag{16}$$

where $\widehat{W}_c^l$ and $\widehat{W}_a^l$ are the NN weights and $\phi_c(x)$ and $\phi_a(x)$ are the associated NN activation functions.

According to Algorithm 3, using the NNs to estimate the solutions will yield the following error:

$$e_c^l = \phi_c^T(x(k))\widehat{W}_c^{l+1} - \sum_{p=k}^{k+\alpha-1} r\left(x(p), \widehat{u}^l(p)\right) - \phi_c^T(x(k+\alpha))\widehat{W}_c^l. \tag{17}$$

Let $\theta = \phi_c^T(x(k))$ and $\sigma^l = \sum_{p=k}^{k+\alpha-1} r(x(p), \widehat{u}^l(p)) + \phi_c^T(x(k+\alpha))\widehat{W}_c^l$. Equation (17) becomes

$$e_c^l = \theta\widehat{W}_c^{l+1} - \sigma^l. \tag{18}$$

To minimize $\|e_c^l\|^2$, we employ the least-square method to update $\widehat{W}_c^{l+1}$. Collect M different data sets for training, where M is a large enough number. Then, one has $\Pi = [\theta_1^T, \theta_2^T, \ldots, \theta_M^T]^T$ and $\Theta^l = [\sigma_1^{lT}, \sigma_2^{lT}, \ldots, \sigma_M^{lT}]^T$. The least-square-based updating law for $\widehat{W}_c^{l+1}$ is given by

$$\widehat{W}_c^{l+1} = \left(\Pi^T\Pi\right)^{-1}\Pi^T\Theta^l. \tag{19}$$

To minimize $r(x(k), \widehat{u}^l(x(k))) + \widehat{V}^l(x(k+1))$, the gradient-descent-based updating law for the actor NN is given by

Let the iteration index $l = 0$.
Select an initial value function $V^0(x)$.
Choose a small enough computation precision ϵ.
Step 2: (Policy Improvement)
With $V^l(x)$, compute the iterative control policy $u^l(x)$ by
$u^l(x(k)) = -(1/2)R^{-1}g^T(x(k))\nabla V^l(x(k+1))$.
Step 3: (Multistep Policy Evaluation)
With $u^l(x)$, calculate the iterative value function $V^{l+1}(x)$ by
$V^{l+1}(x(k)) = \sum_{p=k}^{k+\alpha-1} r(x(p), u^l(p)) + V^l(x(k+\alpha))$.
Step 4: if $\|V^{l+1} - V^l\| \le \varepsilon$, stop and the optimal control policy $u^l(x)$ is acquired; Else, let $l = l + 1$ and go back to Step 2.

Algorithm 3: Multistep HDP.

$$\begin{aligned}\widehat{W}_a^{l,q+1} &= \widehat{W}_a^{l,q} - \beta_a \frac{\partial\left(r\left(x(k), \widehat{u}^{l,q}(x(k))\right) + \widehat{V}^l(x(k+1))\right)}{\partial \widehat{u}^{l,q}(x(k))} \frac{\partial \widehat{u}^{l,q}(x(k))}{\partial \widehat{W}_a^{l,q}} \\ &= \widehat{W}_a^{l,q} - \beta_a \phi_a(x(k)) \left[2R\widehat{u}^{l,q}(x(k)) + g^T(x(k)) \frac{\partial \phi_c(x(k+1))}{\partial x(k+1)} \widehat{W}_c^l\right].\end{aligned} \tag{20}$$

Remark 4. From Table 1 and Figure 3, we can see the performance comparison and relationship among PI algorithm, VI algorithm, and multistep HDP. Due to the existence of initial admissible control, the PI algorithm gets fast convergence. However, the condition of initial admissible control is difficult to realize. Different from the PI algorithm, the initial condition of VI algorithm is easy-to-realize. However, the initial condition may not be admissible, which may lead to the low stability. Multistep HDP follows the initial condition of VI algorithm and develops the multistep policy evaluation step to obtain more history data. Therefore, multistep HDP is easy-to-realize and achieves fast convergence at the same time; that is, multistep HDP successfully combines the advantages of PI and VI algorithms.

5. Application to a Benchmark Power System

The benchmark power system investigated in this paper is illustrated in Figure 4. This power system can be regarded as a microgrid, which is composed of nonpolluting energy (subsystems I and II), load demand sides (subsystem III), and regular generations (subsystem IV). The core control unit is the management center, which maintains the frequency stability against load variations.

5.1. System Model and Application. In Figure 5, first, the real-world power system can be formulated by a state space function via mathematical modeling. After sampling and discretization, the system model can be controlled by computers. Through iterative ADP learning, the approximate optimal control can be obtained. Substituting the approximate optimal control into the system model will yield simulation results. To test the effectiveness of the proposed DT optimal control strategy, let us consider the following power system [19, 20]:

$$\begin{aligned}\Delta\dot{\xi}_f &= -\frac{1}{T_p}\Delta\xi_f + \frac{\alpha_p}{T_p}\Delta\xi_t, \\ \Delta\dot{\xi}_t &= -\frac{1}{T_t}\Delta\xi_t + \frac{1}{T_t}\Delta\xi_g, \\ \Delta\dot{\xi}_g &= -\frac{1}{\alpha_s T_g}\Delta\xi_f - \frac{1}{T_g}\Delta\xi_g + \frac{1}{T_g}u,\end{aligned} \tag{21}$$

where $\Delta\xi_f$ is the frequency deviation; $\Delta\xi_t$ denotes the turbine power; $\Delta\xi_g$ represents the governor position value; T_t, T_g, and T_p denote the time constants of turbine, governor, and power system, respectively; α_p represents the gain of power system; α_s is the speed regulation coefficient; u denotes the control input; and x is the state variable. Let $x = [\Delta\psi_f, \Delta\psi_t, \Delta\psi_g]^T$, where $x_1 = \Delta\psi_f, x_2 = \Delta\psi_t$, and $x_3 = \Delta\psi_g$. Then, the system (21) can be discretized as the form of (1). Set the matrices in the performance index function: $Q = 2I_3$ and $R = 1$.

5.2. Simulation Results. Simulation results are shown in Figure 6. Figure 6(a) implies the system states cannot be stabilized without control. Then, we apply the optimal control strategy into the system. Figure 6(b) indicates the system states can be stabilized after 8 time steps under optimal control. Comparing the trajectories of the system states, the superior control performance of optimal control strategy can be observed. Figure 6(c) shows the 2D plot of convergence trajectory in detail. Figure 6(d) provides the evolution of the control input. The aforementioned simulation results demonstrate the high stability, fast convergence, and low control cost of the DT optimal control strategy.

Table 1: Performance comparison among PI algorithm, VI algorithm, and multistep HDP.

Methods and performance	Computation stability	Initial difficulty	Convergence speed
PI algorithm	High	Difficult	High
VI algorithm	Low	Easy	Low
Multistep HDP	Medium	Easy	Medium

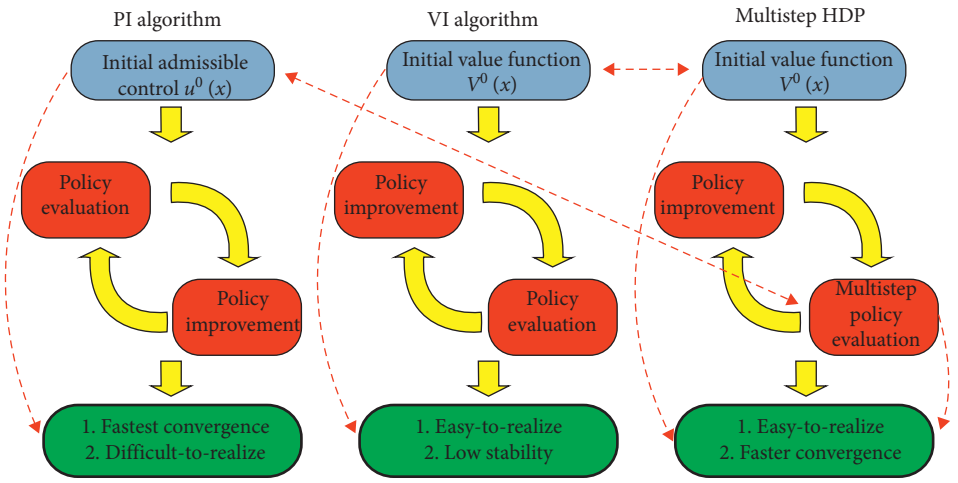


Figure 3: Relationship and difference among PI algorithm, VI algorithm, and multistep HDP.

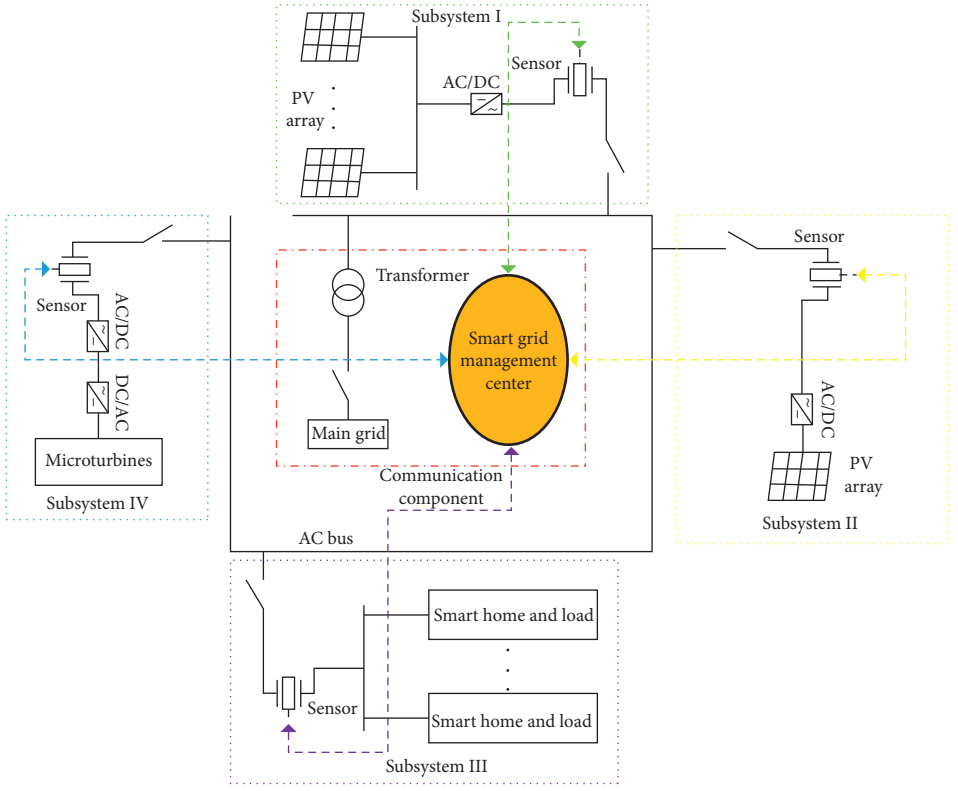


Figure 4: The benchmark power system considered in this paper.

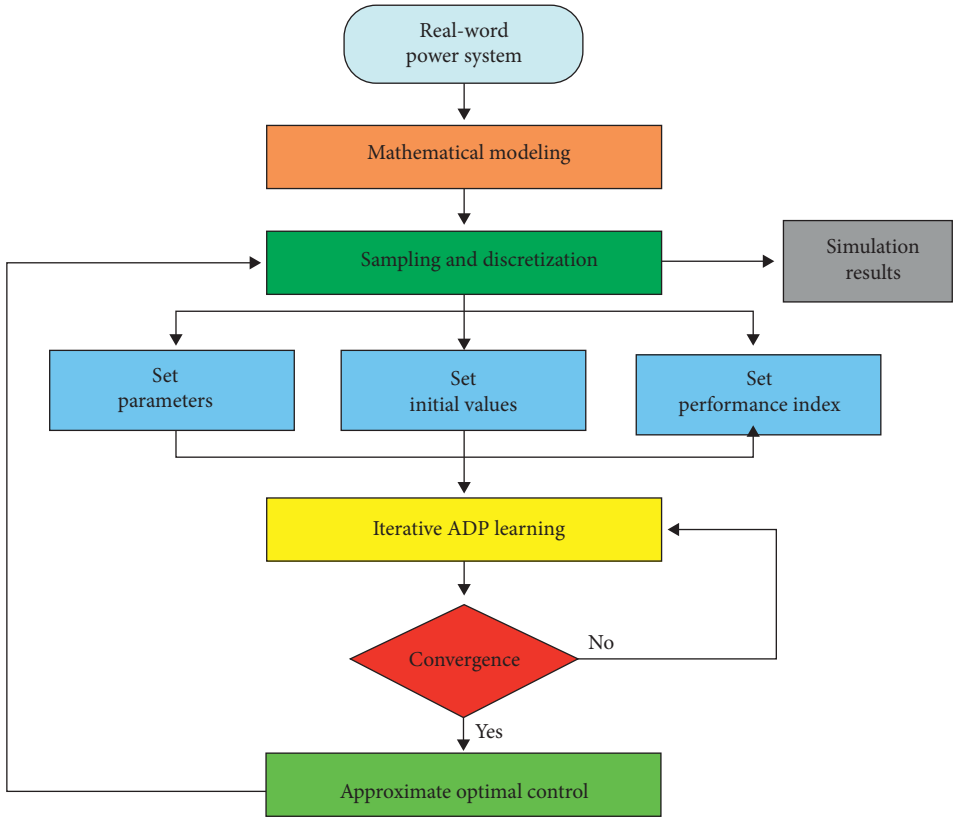
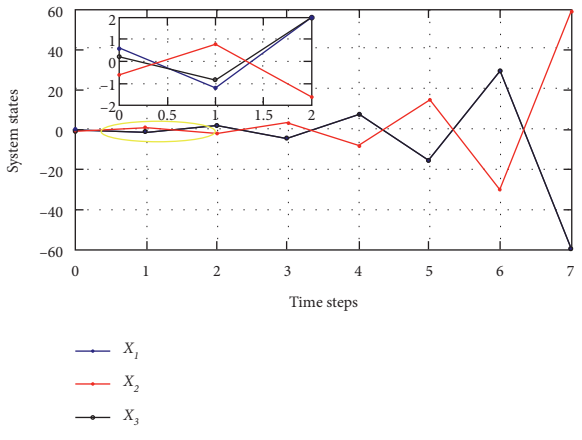
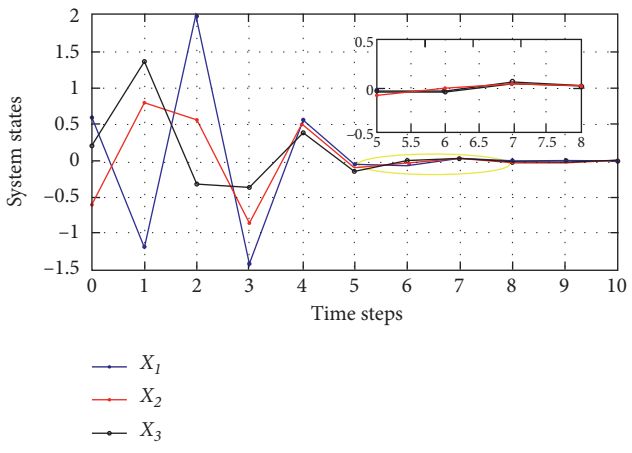


Figure 5: Application to the benchmark power system.

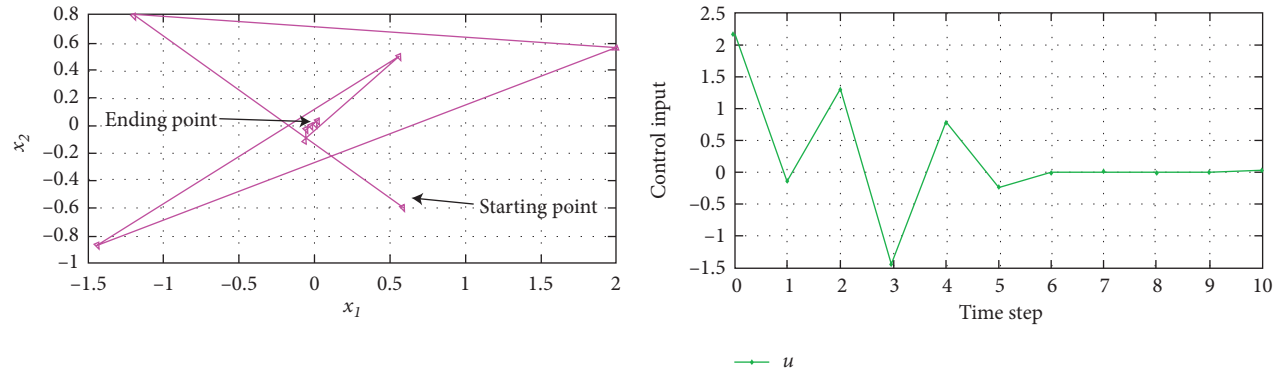


(a)



(b)

Figure 6: Continued.



(c) (d)

FIGURE 6: Simulation results: (a) system states without control; (b) system states with optimal control; (c) 2D plot of x_1 and x_2; (d) control input.

6. Conclusions

In this paper, several state-of-the-art ADP-based methods have been reviewed to address the optimal control problem of DT systems. A comprehensive comparison has been made between PI and VI. A novel multistep HDP method has been introduced to integrate the advantages of PI and VI algorithms with either strict requirement of initial admissible control or longer interaction learning process. The simulation results have demonstrated the effectiveness of our proposed schemes.

Data Availability

Data are available upon request to the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Science and Technology Foundation of SGCC (Grant no. SGLNDK00DWJS1900036).

References

[1] Y. Yang, C. Xu, D. Yue, X. Zhong, X. Si, and J. Tan, "Event-triggered ADP control of a class of non-affine continuous-time nonlinear systems using output information," *Neurocomputing*, vol. 378, pp. 304–314, 2020.

[2] D. Liu, Q. Wei, and P. Yan, "Generalized policy iteration adaptive dynamic programming for discrete-time nonlinear systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 12, pp. 1577–1591, 2015.

[3] R. Wang, Q. Sun, W. Hu, Y. Li, D. Ma, and P. Wang, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, pp. 7866–7879, 2021.

[4] B. Luo, D. Liu, H.-N. Wu, D. Wang, and F. L. Lewis, "Policy gradient adaptive dynamic programming for data-based optimal control," *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3341–3354, 2017.

[5] R. Wang, Q. Sun, D. Ma, and Z. Liu, "The small-signal stability analysis of the droop-controlled converter in electromagnetic timescale," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1459–1469, 2019.

[6] D. Liu, X. Yang, D. Wang, and Q. Wei, "Reinforcement-learning-based robust controller design for continuous-time uncertain nonlinear systems subject to input constraints," *IEEE Transactions on Cybernetics*, vol. 45, no. 7, pp. 1372–1385, 2015.

[7] R. Wang, Q. Sun, D. Ma, and X. Hu, "Line impedance cooperative stability region identification method for grid-tied inverters under weak grids," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 2856–2866, 2020.

[8] D. Zhao and Y. Zhu, "MEC-a near-optimal online reinforcement learning algorithm for continuous deterministic systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 2, pp. 346–356, 2015.

[9] Q. Zhang, D. Zhao, and D. Wang, "Event-Based robust control for uncertain nonlinear systems using adaptive dynamic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 1, pp. 37–50, 2018.

[10] Y. Wang, J. Sun, H. He, and C. Sun, "Deterministic policy gradient with integral compensator for robust quadrotor control," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 10, pp. 3713–3725, 2020.

[11] X. Yang and H. He, "Event-triggered robust stabilization of nonlinear input-constrained systems using single network adaptive critic designs," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 9, pp. 3145–3157, 2020.

[12] B. Luo, D. Liu, and H.-N. Wu, "Adaptive constrained optimal control design for data-based nonlinear discrete-time systems with critic-only structure," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2099–2111, 2018.

[13] W. Zhou, H. Liu, H. He, J. Yi, and T. Li, "Neuro-Optimal tracking control for continuous stirred tank reactor with input constraints," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4516–4524, 2019.

[14] H. Zhang, X. Cui, Y. Luo, and H. Jiang, "Finite-horizon $H_{\infty }$ tracking control for unknown nonlinear systems with saturating actuators," *IEEE Transactions on Neural*

Networks and Learning Systems, vol. 29, no. 4, pp. 1200–1212, 2018.

[15] Q. Wei, D. Liu, Q. Lin, and R. Song, "Adaptive dynamic programming for discrete-time zero-sum games," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 4, pp. 957–969, 2018.

[16] R. Song, F. L. Lewis, and Q. Wei, "Off-policy integral reinforcement learning method to solve nonlinear continuous-time multiplayer nonzero-sum games," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 704–713, 2017.

[17] Q. Wei and D. Liu, "Data-driven neuro-optimal temperature control of water-gas shift reaction using stable iterative adaptive dynamic programming," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6399–6408, 2014.

[18] Q. Wei, D. Liu, F. L. Lewis, Y. Liu, and J. Zhang, "Mixed iterative adaptive dynamic programming for optimal battery energy control in smart residential microgrids," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 4110–4120, 2017.

[19] D. Wang, H. He, C. Mu, and D. Liu, "Intelligent critic control with disturbance attenuation for affine dynamics including an application to a microgrid system," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 4935–4944, 2017.

[20] X. Yang, H. He, and X. Zhong, "Adaptive dynamic programming for robust regulation and its application to power systems," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5722–5732, 2018.

[21] B. Luo, H.-N. Wu, and T. Huang, "Optimal output regulation for model-free quanser helicopter with multistep q-learning," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 6, pp. 4953–4961, 2018.

[22] D. Wang and X. Zhong, "Advanced policy learning near-optimal regulation," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 3, pp. 743–749, 2019.

[23] Q. Wei, D. Liu, Q. Lin, and R. Song, "Discrete-time optimal control via local policy iteration adaptive dynamic programming," *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3367–3379, 017.

[24] Q. Wei, D. Liu, and H. Lin, "Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems," *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 840–853, 2016.

[25] Q. Wei, F. L. Lewis, D. Liu, R. Song, and H. Lin, "Discrete-time local value iteration adaptive dynamic programming: convergence analysis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 6, pp. 875–891, 2018.

[26] Q. Wei, D. Liu, and Q. Lin, "Discrete-time local value iteration adaptive dynamic programming: admissibility and termination analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 11, pp. 2490–2502, 2017.

[27] B. Luo, D. Liu, T. Huang, X. Yang, and H. Ma, "Multi-step heuristic dynamic programming for optimal control of nonlinear discrete-time systems," *Information Sciences*, vol. 411, pp. 66–83, 2017.

[28] D. Liu, D. Wang, D. Zhao, Q. Wei, and N. Jin, "Neural-network-based optimal control for a class of unknown discrete-time nonlinear systems using globalized dual heuristic programming," *IEEE Transactions on Automation Science and Engineering*, vol. 9, no. 3, pp. 628–634, 2012.

Hindawi
Complexity
Volume 2021, Article ID 5571539, 7 pages
https://doi.org/10.1155/2021/5571539

Research Article

A Short-Term Load Forecasting Model of LSTM Neural Network considering Demand Response

Xifeng Guo, Qiannan Zhao, Shoujin Wang, Dan Shan, and Wei Gong

Information & Control Engineering Faculty, Shenyang Jianzhu University, Shenyang 110168, China

Correspondence should be addressed to Qiannan Zhao; n18941643433@sina.cn

Received 11 January 2021; Revised 18 March 2021; Accepted 7 April 2021; Published 16 April 2021

Academic Editor: Qiuye Sun

As one of the key technologies for accelerating the construction of the ubiquitous Internet of Things, demand response (DR) not only guides users to participate in power market operations but also increases the randomness of grid operations and the difficulty of load forecasting. In order to solve the problem of rough feature engineering processing and low prediction accuracy, a short-term load forecasting model of LSTM neural network considering demand response is proposed. First of all, in view of the strong randomness and complexity of input features, the weighted method is used to process multiple input features to strengthen the contribution of effective features and tap the potential value of features. Secondly, an improved genetic algorithm (IGA) is used to obtain the best LSTM parameters; finally, the special gate structure of the LSTM model is used to selectively control the influence of input variables on the model parameters and perform load forecasting. The experimental results show that the research has high prediction accuracy and application value and provides a new way for the development of power load forecasting.

1. Introduction

In recent years, due to the shortage of fossil energy such as oil and the serious environmental problems caused by global carbon emissions, more and more experts and scholars focus on the direction of "Energy Internet [1–3]." People aim to establish a more flexible and stable form of hybrid power generation. As an important constraint for the safe and stable operation of power system, accurate power load forecasting plays an important role in the economic, safe, and reliable operation of the power system.

Demand response (DR) serves as important means to shift energy supply and uses across time to counter the indeterminate variations. It is of great significance to improve the overall energy efficiency of the energy system [4]. However, especially in the Regional Energy Internet System [5], the strong uncertainty of the user-side DR will increase the difficulty of load forecasting. Accurate load forecasting is the basis for realizing the economic operation and scientific management of the power system, as well as the inevitable choice for long-term strategic planning and operational decision-making. In fact, the fundamental requirement for the implementation of DR programs is the load forecasting of its participants whether they are the group of residential consumers managed by aggregators, industrial consumers, or commercial consumers. DR capacity and DR events to be involved in these programs can be determined through load forecasting [6]. Instantly, the relevant technologies of artificial intelligence (AI) and machine learning (ML) have been widely used in renewable energy forecasting, load forecasting, power system fault diagnosis and optimal dispatching, power grid data visualization, and so forth [7]. At the same time, with the installation and application of a large number of measurement equipment such as smart meters [8] and weather monitoring [9], power grid companies have obtained unprecedented large amounts of data. It provides a lot of data support for power load forecasting based on AI [10]. On this basis, to improve the accuracy of load forecasting and the generalization ability of the load forecasting model has become an urgent problem to be solved.

Load forecasting is one of the most widely used areas of artificial intelligence technology in power systems. Scholars have carried out extensive research on the theory and methods of load forecasting. The methods for power load

forecasting mainly include traditional methods and artificial intelligence methods. With the development of the times, traditional methods [11–13] have gradually exposed problems such as poor model generalization ability, difficulty in determining model structure, and difficulty in selecting parameters. Artificial intelligence methods [14–18] have gradually occupied the stage of load forecasting. Ahmad et al. [19] used two ML algorithms to forecast regional short-term energy demand. In [20], deep recurrent neural network (DRNN) and deep convolutional neural network (DCNN) were presented for day-ahead multistep load forecasting in commercial buildings. Khan et al. used grey correlation based on random forest (RF) and mutual information (MI) for feature selection. Kernel principle component analysis (KPCA) was used for feature extraction and enhanced convolutional neural network (CNN) was used for classification [21]. The choice of model hyperparameters affects the entire prediction process. The essence of optimal model configuration is actually a combination problem. Bouktif et al. used metaheuristic-search-based algorithms to reduce the complexity of the search. This method can find the optimal LSTM hyperparameter set more accurately and quickly [22]. For the problem of hyperparameters, Santra and Lin have similar views. Santra and Lin proposed using GA to optimize the initialization parameters of LSTM. This method improved the robustness of short-term load forecasting [23]. Rong and de León [24] proposed a load estimation method suitable for complex power networks. In order to solve the problem of network unobservability, Rong and de León proposed a nonlinear power-temperature curve to predict the load that varies with temperature. The "effective temperature" was introduced in the paper to properly consider the impact on the power consumption of heating and cooling equipment. This method has been well verified in practice. Li et al. [25] proposed a short-term load forecasting method that considers demand response under an energy Internet environment. The grey correlation analysis method was used to process meteorological data to obtain similar daily characteristic variables input to the forecast model. The LSTM neural network model was used for load forecasting. This method separates the DR electricity price that causes users to participate in DR to indirectly consider the DR load. This method provides a new idea for load forecasting considering DR. The rapid development of AI provides more possibilities for load forecasting technology. At the same time, it also puts forward stricter requirements on the quality of load forecasting, which brings new challenges and opportunities to load forecasting technology.

In order to improve prediction accuracy and model generalization ability, a short-term load forecasting model of LSTM neural network considering DR is proposed in this paper. Based on characteristics of engineering processing, the weighted method [26] is used to deal with multiple input features. In order to strengthen the contribution of effective features and explore the potential value of effective features in depth, this paper defines the weights of input features. Then, IGA is used to select the optimal model parameters of LSTM [27]. Finally, the parameter-tuned LSTM model is used for load forecasting.

2. Feature Extraction

This paper uses the data of New South Wales, Australia, from 2006 to 2010 for experimental verification. The data from 2006 to 2008 is used as the training set and the data from 2009 to 2010 is used as the verification set. In this paper, the maximum temperature, the minimum temperature, the average temperature, the date, and the real-time electricity price of the area are extracted as the input features.

2.1. Extraction of Temperature Characteristics. The average temperature is the true average temperature of each day. The average temperature in winter is the lowest and the average temperature in summer is the highest. The load in these two quarters is higher than in other quarters due to the increase in air conditioner usage. Therefore, the feature of average temperature can reduce the abnormal high temperature and low temperature to a reasonable range to reduce temperature fluctuation and then combine it with the real temperature. In this paper, the average temperature weight is set to 0.6, and the true temperature weight is set to 0.4. The specific formula is as follows:

$$T = 0.6 * T_{\text{avg}} + 0.4 * T_{\text{true}}. \tag{1}$$

T is the input feature of temperature, T_{avg} is the average temperature, and T_{true} is the real temperature. Figure 1 shows the temperature characteristic curve after weighting treatment.

2.2. Extraction of Date Characteristics. Date type is another important influencing factor of short-term forecasting. The load on the user side on weekends will be greater than the load on workdays, and the load on the user side on holidays will also be greater than the load on workdays. This paper lists the date type as one of the factors that affect load forecasting. Considering that the parameters of the deep neural network need to be fed back, this paper expresses workdays as 1 and weekends and holidays as 2. Store the date feature in the fourth column of the feature vector.

2.3. Extraction of Real-Time Electricity Price Characteristics. The reform of the electricity market is an inevitable trend and requirement of national development. The marketization of the electricity prices is the top priority, and real-time electricity price is an important factor affecting load fluctuation. First, the maximum information coefficient (MIC) [28] is used to analyze the correlation between load and real-time point price. MIC is used to measure the degree of linear or nonlinear dependence between two variables. The greater the mutual information between the two variables is, the stronger the correlation is. MIC can overcome the shortcoming that mutual information is inconvenient to calculate continuous variables and can better reflect the degree of correlation between attributes. For a binary data set $D \in R^2$ divide D into a grid of x columns and y rows. For the divided grid G, calculate the probability of each unit in G

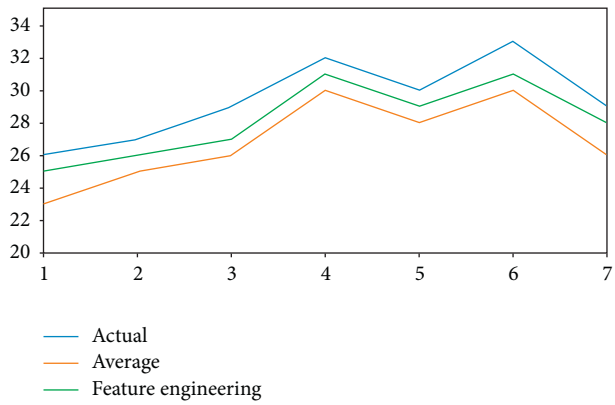


Figure 1: Feature engineering.

to obtain the probability distribution $D_{|G}$ of the binary data set D on the grid G. Save it as $I^*[D(x, y)]$:

$$I^*[D(x, y)] = \max I[D(x, y)]. \tag{2}$$

Standardize the obtained mutual information and find the maximum mutual information coefficient as follows:

$$M(D)_{x,y} = \frac{I^*[D(x, y)]}{\log \min\{x, y\}}, \tag{3}$$

$$s.t.\ xy < B(n), \tag{4}$$

$$F(D)_{\mathrm{MIC}} = \max_{xy<B(n)}\left\{M(D)_{x,y}\right\}, \tag{5}$$

where n is the sample size; $B(n)$ is a function of the sample size, which represents the constraint of the total number of grids xy in grid G, which must be less than $B(n)$; and generally $B(n) = n\hat{0}.6$. In essence, MIC is a normalized maximum mutual information, with a value interval of $[0, 1]$. The statistical results show that MIC > 0.4 days, account for 908 days in total, accounting for 46.69%, of which MIC > 0.6 days account for 502 days, accounting for 26.43%.

Figure 2 is the curve of electricity price and load. It can be seen that the trend of electricity price and load is basically the same. Residents can check the real-time electricity price to appropriately reduce electricity consumption and reduce the load, which will play a role in peak clipping. When the load reaches a trough, the electricity price will also be reduced, so as to encourage users to increase the amount of electricity to fill the valley. It has been proved by experiments that the electricity price has a great correlation with the load. In this paper, the real-time price is put into the fifth column of the feature vector as the input feature.

3. Prediction Model

3.1. Improved Genetic Algorithm. Genetic algorithm (GA) is a global search optimization algorithm formed according to the selection and genetic law of "survival of the fittest" in nature. The initial population composed of feasible solutions is operated by three operators of selection, crossover, and mutation. Eliminate one by one according to the fitness value of the individual. Finally, the individuals with the best fitness are left to form a new population, parallel calculation of multiple solution groups by genetic algorithm. Replace generations to achieve global convergence to obtain the optimal solution.

Hill climbing is a good local search algorithm. Firstly, a point is randomly selected as the initial point of iteration in the search space, and then a point is randomly generated in its neighborhood to calculate its function value. If the function value of the point is better than the current point, the current point is used to replace the initial point as a new initial point to continue searching in the neighborhood. Otherwise, the search process will be terminated if another point and initial point are randomly generated in the neighborhood until the better point is found or the better point is not found several times in a row. Because hill climbing is a method to optimize by randomly generating individuals in the neighborhood, it does not need to use a gradient. Hill climbing method can play the role of local optimization when genetic algorithm deals with complex problems. In the iterative process of this paper, the hill climbing method is introduced to optimize the individuals obtained by genetic algorithm, which greatly improves the efficiency of optimization.

3.2. LSTM Algorithm Mechanism. LSTM is an excellent variant of Recurrent Neural Network (RNN), which solves the problem of short memory and difficult training of RNN. It is very suitable for the classification and prediction of time series and is widely used in natural language processing. As shown in Figure 3, t is the time point; i_t is the input gate, which is mainly used to control how much information about the current state of the network needs to be saved in the internal state; f_t is the forget gate, which is mainly used to control how much information is in the past state and needs to be discarded; o_t is the output gate, mainly used to control how much information of the internal state at the current moment needs to be output to the external state; h_t is the internal state of the neuron at the current moment; c_t is the external state at the current moment; x_t is the external input at the current moment; and σ is the activation layer function. The calculation formula between each variable is as follows:

$$i_t = \sigma\left[W_{hi} \times h_{t-1} + W_{xi} \times x_t + W_{ci} \times c_{t-1} + b_i\right], \tag{6}$$

$$f_t = \sigma\left[W_{hf} \times h_{t-1} + W_{xf} \times x_t + W_{cf} \times c_{t-1} + b_f\right], \tag{7}$$

$$o_t = \sigma\left[W_{ho} \times h_{t-1} + W_{xo} \times x_t + W_{co} \times c_{t-1} + b_o\right], \tag{8}$$

$$c_t = f_t c_{t-1} + i_t \tanh\left[W_{hg} \times h_{t-1} + W_{xc} \times x_t + b_c\right], \tag{9}$$

$$h_t = o_t \tanh\left[c_t\right]. \tag{10}$$

Through the comprehensive application of these three control gates, the LSTM neural network can control the retention and discarding of the information transmitted in the neural network. It determines how much of the new state information at the current moment needs to be stored in the

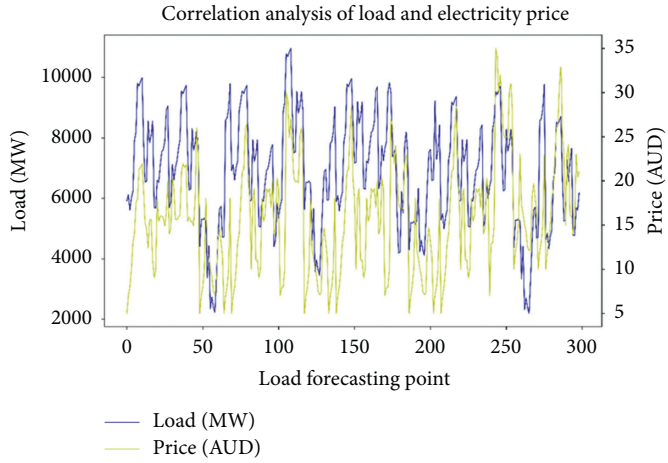


Figure 2: Curve of electricity price and load.

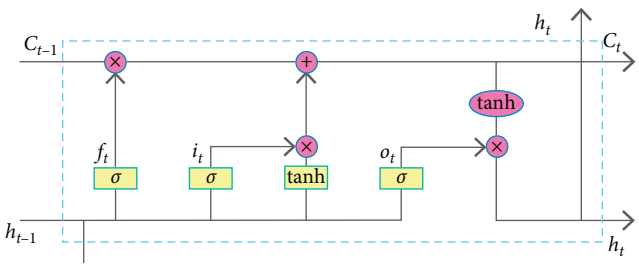


Figure 3: Model of LSTM.

memory unit. Compared with general neural networks, LSTM neural networks can learn dependencies with a relatively long span without the problems of gradient disappearance and gradient explosion [29].

3.3. Selection of LSTM Optimal Parameters. The initial values of the weighting matrices can affect the performance of the LSTM. Inspired by [23], an improved genetic algorithm (IGA) is used to assist in searching the proper initial values for the weighting matrices of the LSTM. Figure 4 shows the flow diagram of the proposed method.

Step 1. Initialize chromosome population.

A set of n chromosomes are randomly generated. Each chromosome w contains the values for all the weighting matrices in LSTM.

Step 2. Set fitness function.

For each chromosome in the current population, its value is used to initialize the weighting matrices in the LSTM. The LSTM predicts the output error of the training sample as the fitness function and calculates the fitness value of the individual in the initial population.

Step 3. Generate the new population through genetic operations.

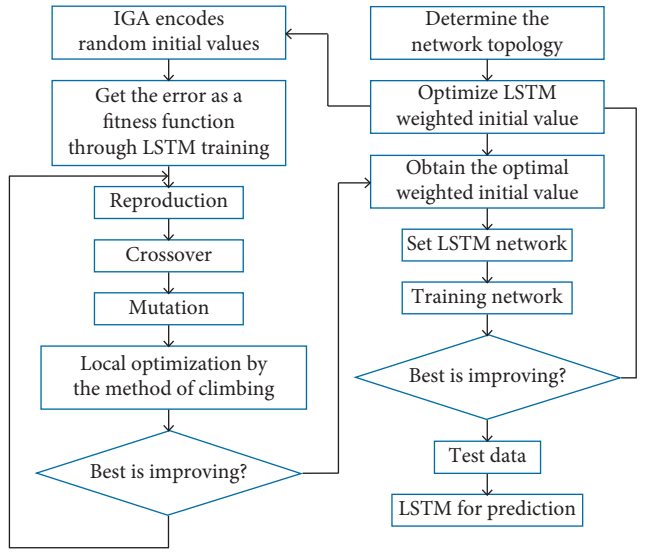


Figure 4: The flow diagram of the proposed method.

This step generates a new population that contains the same number of chromosomes as the current population. The roulette wheel selection is used so that a chromosome with a higher fitness value had a higher probability of being selected for genetic operations. The chromosomes of the new population are produced by reproduction, crossover, and mutation operations on the chromosomes selected from the current population.

Chromosomes with higher fitness in the current population will be selected for reproduction, and it will be added to the new population, until the reproduction ratio is reached. The elitism policy is used in this article to ensure that the best chromosome among the current chromosomes can be added to the new population group.

Repeatedly select two chromosomes of the current population as the parent chromosomes for the crossover operation to generate two offspring chromosomes and add them to the new population until the crossover ratio is

reached. Uniform crossover is adopted in this study. With uniform crossover, each value in the offspring chromosomes is independently chosen from the two values at the same corresponding position in the two parent chromosomes, as shown in Figure 5.

The mutation operation repeatedly selected a chromosome from the current population, modified the selected chromosome to generate a new chromosome, and added the mutant to the new population, until the mutation ratio has reached. With one-point mutation, a small random change is injected into the value of a randomly selected position in the selected chromosome to generate the mutant, as shown in Figure 6. Mutation operation makes the chromosome population more diverse.

Step 4. Select the best chromosome.

After one evolution, the fitness value is calculated for the retained chromosomes, and the best chromosome is retained. Compare the fitness value of the chromosome with the maximum fitness value and the best chromosome of the previous generation. The best chromosome is recorded. Finally, the chromosome with the best fitness value is used to initialize the weighting matrix in the LSTM, and the training data is fed to the LSTM to generate the fitness value of the chromosome.

Step 5. Stop criterion

If the fitness value of the best chromosome is already optimal, then the LSTM of the best chromosome is adopted, and the proposed method is terminated. The parameters of LSTM are initialized with the optimal weights to obtain the optimal network structure.

In this paper, the original real data set is divided into three training sets and a test set. Feature engineering processing is carried out by means of the weighted discretization method. The input features processed by the weighting method are more valuable, and then the IGA is used to select the optimal parameters of the LSTM network. The LSTM network after parameter tuning is used to train and predict real data. The specific process is shown in Figure 7.

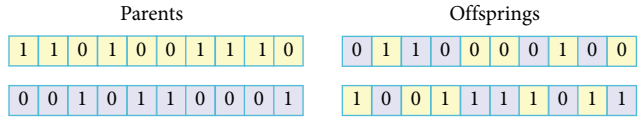


Figure 5: Uniform crossover.

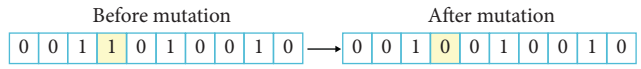


Figure 6: One-point mutation.

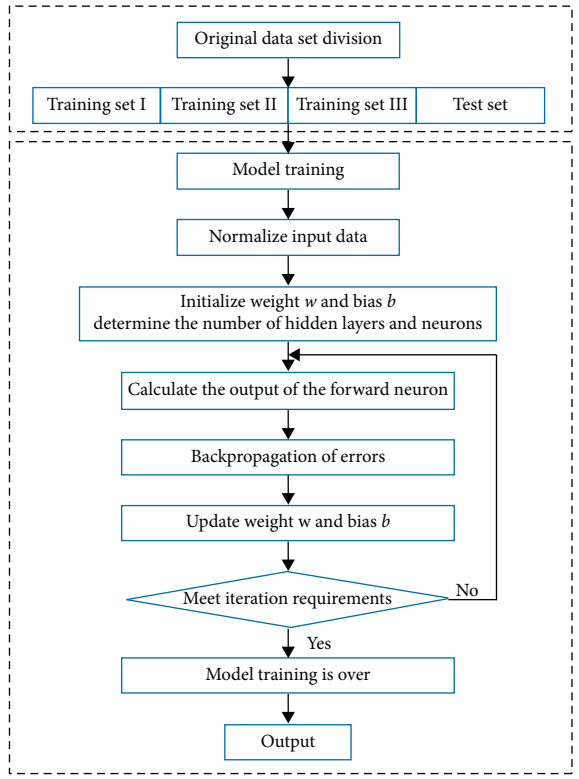


Figure 7: Proposed model.

4. Case Study

In order to evaluate the prediction performance, the mean absolute percentage error Y_{MAPE} and forecasting accuracy Y_{FA} are set according to the load forecasting indexes evaluated by State Grid Corporation of China. The specific formulas are as follows:

$$Y_{\mathrm{MAPE}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{X_{\mathrm{act}}(i) - X_{\mathrm{pred}}(i)}{X_{\mathrm{act}}(i)}\right|, \tag{11}$$

$$Y_{\mathrm{FA}} = \left(1 - \frac{\left|X_{\mathrm{act}}(i) - X_{\mathrm{pred}}(i)\right|}{X_{\mathrm{act}}(i)}\right) \times 100\%, \tag{12}$$

where n is the total number of predictions and $X_{\mathrm{act}}(i)$ and $X_{\mathrm{pred}}(i)$ are the real value and the predicted value of the load at time i respectively.

4.1. Analysis of Experimental Results. After training and verification of the proposed model, this paper conducted day-ahead short-term load forecasting from 2009 to 2010 and calculated the Y_{MAPE} and the Y_{FA} of the parameter-tuned SVM model and RF model. The results are shown in Table 1. It can be seen from the table that whether it is a one-day or half a year, Y_{MAPE} of the proposed model is the smallest and Y_{FA} is the greatest and the goodness of fit is the best of all models, which shows that the method proposed in this paper is more accurate and the robustness is better.

Experimental verification shows that the comparison between the four methods and the real load curve is shown in Figure 8. It can be seen from the figure that the load curve predicted by the model proposed in this paper has the highest degree of fit with the real curve. The prediction results are more accurate than the other three methods. In

Table 1: Comparison of model prediction results.

Prediction model	One day		Half a year	
	Y_{MAPE}	Y_{FA} (%)	Y_{MAPE}	Y_{FA} (%)
Proposed	0.1067	89.31	0.0901	90.99
SVM	0.2025	79.75	0.1862	81.38
RF	0.1689	83.11	0.1259	87.41
GRU	0.1219	87.81	0.1015	89.85

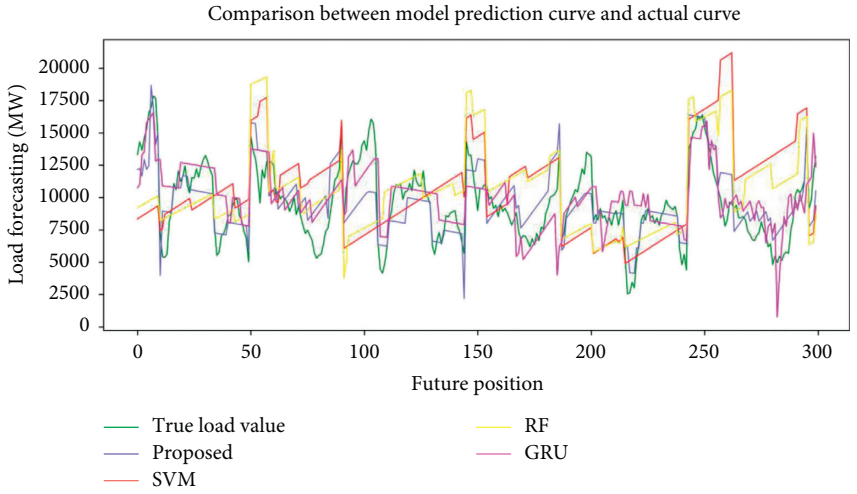


Figure 8: Model prediction comparison diagram.

order to further evaluate the performance of the model, this article compares the calculation speed of each model. The average calculation time of this model is 7.96 s, while the average calculation speeds of SVM, RF, and GRU are 11.14 S, 9.86 S, and 8.03 s, respectively. Through the comparison of average calculation time, it can also be seen that the model proposed in this article has certain advantages compared with other models.

5. Conclusions

In view of the phenomenon that user-side DR is highly uncertain and load prediction becomes more difficult, this paper selects regions with relatively perfect demand-side management as target regions and puts forward a short-term load prediction method that takes DR into consideration. In this paper, the weighted discretization method is used to process the input features and enhance the contribution degree of input features. Then, IGA is used to select the optimal model parameters of LSTM. After that, the LSTM prediction model is established for load prediction. Finally, the effectiveness and superiority of the proposed method are verified based on the actual data, and the following conclusions are drawn.

(1) Considering a variety of load influencing factors, this paper uses the weighted discretization method to process input characteristics. And the maximum information coefficient method is used to verify the real-time electricity price characteristics selected in this paper have a strong correlation with load.

(2) In this paper, the IGA algorithm is used to select the optimal LSTM parameters.

(3) The LSTM model has a special forgetting gate mechanism, which can automatically screen the input variables that are beneficial to the model during the training process. It cuts down the prediction time on the basis of improving the model performance and prediction accuracy.

Data Availability

The data used in the study are available at http://www.aemo.com.au.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by National Key Research and Development Plan (2019YFF0301505); Liaoning Natural Science Foundation (2019-ZD-0682); and Education Department of Liaoning Province Science and Technology Research Projects (LNJC201912).

References

[1] R. Wang, Q. Sun, and W. Hu, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *IEEE Transactions on Power Electronics*, vol. 99, p. 1, 2021.

[2] D. Ma, X. Hu, and H. Zhang, "A hierarchical event detection method based on spectral theory of multidimensional matrix for power system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 99, pp. 1–14, 2019.

[3] Q. Sun, R. Han, H. Zhang, J. Zhou, and J. M. Guerrero, "A multiagent-based consensus algorithm for distributed coordinated control of distributed generators in the energy internet," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 3006–3019, 2015.

[4] H. Xiao, W. Pei, and Z. Dong, "Bi-level planning for integrated energy systems incorporating demand response and energy storage under uncertain environments using novel metamodel," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 2, 2018.

[5] N. Li and K. Zheng, "Regional energy internet construction framework and key tasks," *IOP Conference Series: Earth and Environmental Science*, vol. 446, p. 22034, 2020.

[6] M. Mansoor, F. Grimaccia, and S. Leva, "Comparison of echo state network and feed-forward neural networks in electrical load forecasting for demand response programs," *Mathematics and Computers in Simulation*, vol. 184, pp. 282–293, 2020.

[7] J. Shi and J. Zhang, "Load forecasting based on multi-model by stacking ensemble learning," *Proceedings of the CSEE*, vol. 39, no. 14, pp. 4032–4041, 2019.

[8] A. M. Alonso, F. J. Nogales, and C. Ruiz, "Hierarchical clustering for smart meter electricity loads based on quantile autocovariances," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4522–4530, 2020.

[9] T. Ahmad, W. A. Shah, and D. Zhang, "Efficient energy planning with decomposition-based evolutionary neural networks," *IEEE Access*, vol. 99, 2020.

[10] X. Kong, C. Li, and F. Zheng, "A short-term load forecasting method based on EMD and characteristic correlation analysis," *Automation of Electric Power Systems*, vol. 43, no. 5, pp. 75–85, 2019.

[11] K. Wan and R. Liu, "Application of interval time-series vector autoregressive model in short-term load forecasting," *Power System Technology*, vol. 36, no. 11, pp. 77-78, 2012.

[12] S. Fan and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," *Monash Econometrics & Business Stats Working Papers*, vol. 27, no. 1, pp. 134–141, 2010.

[13] E. Paparoditis and T. Sapatinas, "Short-term load forecasting: the similar shape functional time-series predictor," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3818–3825, 2013.

[14] Z. Wu, X. Kong, and X. Dong, "Short-term spatial load forecasting based on partition and classification of power distribution information," *Proceedings of the CSU-EPSA*, vol. 31, no. 2, pp. 30–35, 2019.

[15] L. Sehovac and K. Grolinger, "Deep learning for load forecasting: sequence to sequence recurrent neural networks with attention," *IEEE Access*, vol. 8, pp. 36411–36426, 2020.

[16] K. Park, S. Yoon, and E. Hwang, "Hybrid load forecasting for mixed-use complex based on the characteristic load decomposition by pilot signals," *IEEE Access*, vol. 7, pp. 12297–12306, 2019.

[17] G. Suo, L. Song, and Y. Dou, "Multi-dimensional short-term load forecasting based on XGBoost and fireworks algorithm," in *Proceedings-2019 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science*, pp. 245–248, DCABES, Wuhan, China, November 2019.

[18] Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu, and Z. Zhu, "Multiscale convolutional neural network with time-cognition for multi-step short-term load forecasting," *IEEE Access*, vol. 7, no. 99, pp. 88058–88071, 2019.

[19] T. Ahmad, H. Chen, and Y. Huang, "Short-term energy prediction for district-level load management using machine learning based approaches," *Energy Procedia*, vol. 158, pp. 3331–3338, 2019.

[20] M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," *Applied Energy*, vol. 236, pp. 1078–1088, 2019.

[21] A. B. M. Khan, N. Javaid, and O. Nazeer, "Hourly electricity load forecasting in smart grid using deep learning techniques," *Innovative Mobile and Internet Services in Ubiquitous Computing*, vol. 19, pp. 185–196, 2020.

[22] S. Bouktif, A. Fiaz, and A. Ouni, "Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting," *Energies*, vol. 13, 2020.

[23] A. S. Santra and J.-L. Lin, "Integrating long short-term memory and genetic algorithm for short-term load forecasting," *Energies*, vol. 12, no. 11, p. 2040, 2019.

[24] H. Rong and F. de León, "Load estimation of complex power networks from transformer measurements and forecasted loads," *Complexity*, vol. 2020, Article ID 2941809, , 2020.

[25] C. Li, X. Kong, and S. Zhu, "Short-term load forecasting of regional power grid considering demand response in energy interconnection environment," *Automation of Electric Power Systems*, vol. 45, pp. 1–11, 2020.

[26] Y. Xu, F. Zuo, and X. Zhu, "Research on load prediction based on improved GBDT algorithm," *Proceedings of the CSU-EPSA*, vol. 32, pp. 1–8, 2020.

[27] F. Lv, Yi Liu, and Y. Qi, "Short-term load forecasting based on optimized learning machine using improved genetic algorithm," *Journal of North China Electric Power University*, vol. 45, no. 6, pp. 1–7, 2018.

[28] P. Wen, J. Wang, and S. Yin, "Short-term load forecasting model based on Attention-LSTM in electricity market," *Power System Technology*, vol. 5, pp. 1745–1751, 2019.

[29] G. B. Goh, N. O. Hodas, and A. Vishnu, "Deep learning for computational chemistry," *Journal of Computational Chemistry*, vol. 38, no. 16, pp. 1291–1307, 2017.

Hindawi
Complexity
Volume 2021, Article ID 5585892, 10 pages
https://doi.org/10.1155/2021/5585892

Research Article

Improved Tomlinson–Harashima Precoding for Ultra Reliable Communication in Intelligent Transportation Systems

Shibiao He,[1] **Xinyi Yang**,[2] **and Yong Liao**[2]

[1]*School of Electronic Information, Chongqing Institute of Engineering, Chongqing 400056, China*
[2]*School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China*

Correspondence should be addressed to Shibiao He; hdoctor@vip.sina.com

Received 23 February 2021; Revised 8 March 2021; Accepted 24 March 2021; Published 5 April 2021

Academic Editor: Renke Han

Cyber-physical systems (CPSs) are characterized by integrating computation and physical processes. To cope with the challenges of the application of the CPSs in all kinds of environments, especially the cellular vehicle-to-everything (C-V2X) which needs high quality end-to-end communication, the robustness and reliability for CPSs are very crucial. Aiming at the technical challenges of information transmission caused by the fading effect and the fast time-varying characteristics of the channel for C-V2X communication, an improved Tomlinson–Harashima precoding (THP) algorithm for multiple input multiple output (MIMO) systems is proposed. Channel state information (CSI) and correlation are exploited to compensate instantaneous CSI, which could reflect current real-time channel status exactly. Further, the iterative water filling power allocation algorithm and the multiuser scheduling algorithm based on the greedy algorithm are jointly optimized and applied to the THP, which could improve the system performance. Simulation results show that the proposed algorithm can be efficiently applied to high-speed mobility scenarios and improve bit error ratio (BER) performance as well as spectrum utilization.

1. Introduction

In recent years, the construction of global smart city is accelerated, which makes cyber-physical systems (CPSs) develop rapidly [1]. Large-scale networked CPSs are deployed to address global needs in areas such as energy [2, 3], transportation [4, 5], power system [6, 7], and auto-control [8, 9]. As an intelligent system integrating computing, communication, and control, any problems may affect the normal operation of the CPSs. Thus, the stability of CPSs which includes system reliability and robustness should be considered and improved. CPSs should respond to the input of the system timely and effectively, especially for cellular vehicle-to-everything (C-V2X) communication scenarios. The vehicle in the C-V2X adopts short distance wireless communication technology to communicate between vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) [10]. Ultra reliable low latency communication (URLLC) is required for C-V2X. However, the computing and storage capabilities of vehicle devices are limited, and it is difficult to meet the constraints of a large number of computing requirements and low latency. Therefore, mobile edge computing (MEC) can be introduced into the C-V2X. The C-V2X based on MEC is to move part of the computing power of the core network to the edge of the C-V2X. Generally, the MEC server is equipped with road side unit (RSU). The vehicle uploads the calculation task to the RSU through the V2I communication, and the MEC server corresponding to the RSU calculates the task. When the MEC server completes the calculation, the calculation result is transmitted from the RSU back to the vehicle [11]. And computing requirements and delay constraints can be effectively solved through the combination of MEC and C-V2X. At the same time, the high-speed mobility of the vehicle makes signal propagation attenuate fast, and the Doppler frequency shift effect is serious. Dynamically changing reflective objects on the road will also deteriorate the channel and affect the received signal strength and quality. In order to solve the abovementioned

problems, multiple input multiple output (MIMO) technology has been applied in the C-V2X [12, 13].

MIMO can suppress channel fading, with two modes of spatial multiplexing and spatial diversity [14, 15]. The spatial diversity can provide reliable communication, and MIMO can obtain diversity gain by using precoding at the transmitter. Precoding is divided into linear precoding and nonlinear precoding. Compared with linear precoding, nonlinear precoding has higher spectrum utilization. Tomlinson–Harashima precoding (THP) is a typical nonlinear precoding algorithm, which can make a good tradeoff between complexity and performance [16].

When the vehicle is moving at high-speed, the channel state information (CSI) changes significantly during a symbol transmission period. Therefore, traditional channel estimation is not accurate. The fast time-varying characteristics of the channel cause the feedback CSI cannot accurately describe the real-time channel [17, 18].

Moreover, if the transmitter adopts the unreasonable transmission power allocation algorithm, it will not be able to obtain communication quality of service (QoS) assurance and better spectrum utilization when the channel has a fast time-varying characteristic [19]. Therefore, the joint power allocation algorithms in the precoding design can effectively improve the transmission reliability and spectrum utilization of the system [20].

Precoding design in high-speed mobile scene has become a research hotspot currently. Zhang et al. [21] proposes a simple precoding and detection scheme that implements precoding by dividing modulated data symbols in the frequency domain into groups and inserting multiple redundant carriers on both sides of each data group. The simulation results show that the scheme can maintain good performance in high-speed scenes, but it does not analyze the influence of the scheme on spectrum utilization. Benmimoune and Massicotte [22] presented a precoding method based on Kerdock code. The method is based on codebook with low complexity, but its overall performance is poor. It can be seen that precoding in high-speed mobile scenes still faces challenges.

Based on the research and analysis, an improved THP algorithm in high-speed mobility scenarios is proposed. The algorithm combines the statistical information with the space-time correlation characteristics of the MIMO channel to construct a dynamic CSI channel model. At the same time, the iterative water filling power allocation algorithm and the multiuser scheduling algorithm based on the greedy algorithm are jointly designed with THP, which can effectively improve the system transmission reliability and spectrum utilization. The main contributions of this paper are as follows:

(1) In order to suppress the channel fading caused by high-speed vehicle movement, we combine MIMO and IoV to construct a MIMO-V2I communication system. Moreover, since the channel in the high-speed mobile environment of vehicles has fast time-varying characteristics, we adopt the dynamic CSI model that combines the statistical information with the space-time correlation characteristics of the channel, which can reflect the current channel characteristics accurately.

(2) Based on the proposed dynamic CSI model, in order to obtain the diversity gain of the MIMO system, an improved THP algorithm is proposed. In the high-speed vehicle moving environment, the algorithm applies the iterative water filling power allocation and the user scheduling based on the greedy algorithm to the THP algorithm for joint optimization design. Firstly, the user scheduling algorithm aims to maximize spectrum utilization and selects the optimal user set according to the channel information of the dynamic CSI model. Secondly, according to the user scheduling result, the channel matrix based on the dynamic CSI model is selected and reconstructed. And the reconstructed channel matrix is applied to the THP algorithm, which can effectively eliminate multiuser interference. The iterative water filling algorithm allocates power to each user according to the channel characteristics of each user in the system, realizing the redistribution of total power, which can effectively improve the transmission efficiency and maximizing the spectrum utilization of the system.

The rest of this paper is organized as follows. Section 2 is the system model. Section 3 proposes an improved THP based on the iterative water filling algorithm in high-speed mobility scenarios, which contains a dynamic CSI model and user scheduling algorithm based on greedy and iterative water filling. Section 4 is the complexity analysis and simulation results of the algorithm. Finally, the conclusion is discussed in Section 5.

2. System Model

Assuming that in a multiuser MIMO-IoV system, the infrastructure has deployed N_T transmission antennas. For simplicity, it is assumed that infrastructure and N vehicles transmit data, and each vehicle is equipped with an antenna $N_R = 1$. The N vehicles data are represented by an N-dimensional vector $\mathbf{a} = [a_1, a_2, \ldots, a_N]^T$ and are taken from an M-QAM constellation with energy $\sigma_a^2 = 2(M-1)/3$.

The multiuser THP system model is shown in Figure 1. It consists of a feedback matrix $\mathbf{B}$, N nonlinear operators $\mathrm{mod}_{\sqrt{M}}(x)$, and a feedforward matrix $\mathbf{F}$ with dimensions $N_T \times N$. The transmitting signal enters the high-speed channel after the precoding and reaches the receiver. For system physics to achieve, $\mathbf{B}$ is strictly lower triangular to allow data precoding recursively while the modulo operator acts independently over the real and imaginary parts of its input according to the following rule:

$$\mathrm{mod}_{\sqrt{M}}(x) = x - 2\sqrt{M} \cdot \left\lfloor \frac{x - \sqrt{M}}{2\sqrt{M}} \right\rfloor, \tag{1}$$

where $\lfloor c \rfloor$ indicates the smallest integer larger than or equal to c. In practice, $\mathrm{mod}_{\sqrt{M}}(x)$ performs a periodic mapping of

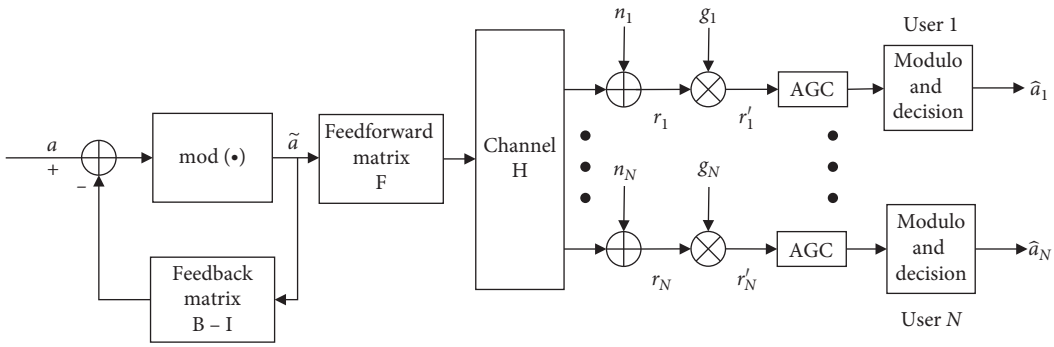


Figure 1: Multiuser THP system model.

the real quantity x onto the interval $(-\sqrt{M}, \sqrt{M}]$. In this way, the precoding symbols $\tilde{a}_n$ are constrained into the square region $\aleph = \{x^{(R)} + jx^{(I)} | x^{(R)}, x^{(I)} \in (-\sqrt{M}, \sqrt{M}]\}$ and the transmit power is correspondingly reduced with respect to linear prefiltering.

According to (1), we see that the precoding symbols $\tilde{a}_n$ can be iteratively computed as follows:

$$\tilde{a}_n = a_n - \sum_{\ell=1}^{n-1} [\mathbf{B} - \mathbf{I}]_{n,\ell} \tilde{a}_l + d_n, \quad n = 1, 2, \ldots, N, \tag{2}$$

where $d_n = 2\sqrt{M} q_n$ and q_n is a complex-valued quantity whose real and imaginary parts are suitable integers that reduce $\tilde{a}_n$ to the square region $\aleph$. Equation (2) indicates that the modulo operator in Figure 1 is equivalent to adding a vector $\mathbf{d} = [d_1, d_2, \ldots, d_N]^T$ to the input data symbols. Since the feedback matrix $\mathbf{B}$ is set to strictly lower triangular, we have defined the modified data vector $\mathbf{v} = \mathbf{a} + \mathbf{b}$. Accordingly, the precoding can be rewritten as

$$\tilde{\mathbf{a}} = \mathbf{v} - (\mathbf{B} - \mathbf{I})\tilde{a} \Longrightarrow \tilde{\mathbf{a}} = \mathbf{B}^{-1}\mathbf{v}. \tag{3}$$

The precoding symbols $\tilde{a}_n$ are fed to the forward matrix $\mathbf{F}$. The resulting N_T-dimensional vector $\mathbf{x} = \mathbf{F}\tilde{a}$ is finally transmitted over the channel using the N_T antennas of the infrastructure. The channel is mathematically described by a $N \times N_T$ dimensional matrix $\mathbf{H}$. In particular, $[\mathbf{H}]_{n,i}$ represents the channel gain from the i-th transmit antenna to the n-th receive antenna. The discrete-time signal at the n-th mobile vehicle can be written as

$$r_n = \mathbf{h}_n \mathbf{F} \tilde{a} + n_n, \tag{4}$$

where $\mathbf{h}_n$ indicates the n-th row of $\mathbf{H}$ while n_n is additive white Gaussian noise with zero-mean and variance σ_n^2. Stacking the received signals of all users into a single vector $\mathbf{r} = [r_1, r_2, \ldots, r_N]^T$ and submitting in (3), it can be expressed as

$$\mathbf{r} = \mathbf{HF}\tilde{\mathbf{a}} + \mathbf{n}, \tag{5}$$

where $\mathbf{n} = [n_1, n_2, \ldots, n_N]^T$ is a Gaussian vector with zero-mean and covariance matrix $\sigma_n^2 \mathbf{I}$. Since the modulo and the feedback operation are performed at the transmitter, it is equivalent to adding a vector $\mathbf{d}$. At the receiver, each sample r_n is passed to the automatic gain control (AGC) and to the same modulo operator employed at the transmitter so as to remove the effect of d_n. Finally, the received vector can be written as

$$\mathbf{r}' = \mathbf{G}(\mathbf{HF}\tilde{a} + \mathbf{n}) = \mathbf{GHFB}^{-1}\mathbf{v} + \mathbf{Gn}. \tag{6}$$

3. Improved THP Based on Iterative Water Filling Algorithm

3.1. Dynamic CSI Based on Statistical Information. In the high-speed mobility scenarios, the channel response h of the wireless channel has the characteristics of time-domain selective fading and fast time-varying under the combined influence of multipath effect and Doppler effect [23]. The channel is modeled according to the geometric model. Considering one scattering, the channel impulse response (CIR) expression of the downlink of the MIMO system can be obtained as [23]

$$h_{pq}(t) = h_{pq}^{\mathrm{LOS}}(t) + h_{pq}^{\mathrm{SB}}(t), \tag{7}$$

where h_{pq}^{LOS} and h_{pq}^{SB} denote the line-of-sight (LOS) and single-bounced (SB) component of the p-th transmit antenna to the q-th receive antenna, respectively.

$$h_{pq}^{\mathrm{LOS}}(t) = Ce^{-j2\pi f_c \tau_{pq}(t)} e^{-j2\pi f_{\max} t \cos\left(\phi_{T_p}^{\mathrm{LOS}} - \gamma_R\right)},$$

$$h_{pq}^{\mathrm{SB}}(t) = C \lim_{N \longrightarrow \infty} \sum_{n=1}^{N} \frac{1}{\sqrt{N}} e^{-j\left(\psi_n - 2\pi f_c \tau_{pq,n}(t)\right)} e^{j2\pi f_{\max} t \cos\left(\phi_R^n(t) - \gamma_R\right)}, \tag{8}$$

where $C = \sqrt{(K_{pq}/K_{pq} + 1)}$, K is the Rice factor, τ denotes the delay on the link, and the link delays on different paths are different, which will result in multipath effects. $e^{-j2\pi f_{\max} t \cos(\phi_{T_p}^{\mathrm{LOS}} - \gamma_R)}$ and $e^{j2\pi f_{\max} t \cos(\phi_R^n(t) - \gamma_R)}$ denote the Doppler shift. $f_{\max}$ is the maximum Doppler shift of the channel, which is determined by the moving speed of the vehicle. Due to the different angles of arrival (AoA) of the beams on different paths, a Doppler spread effect will be produced, resulting in time-domain selective fading. The coherence time is inversely proportional to the Doppler spread. The larger the Doppler shift, the shorter the coherence time. When the coherence time is less than a symbol transmission period, the current transmission signal

experiences fast fading. When the channel has a fast time-varying characteristic, the CSI changes significantly. According to the above channel model, the channel matrix can be expressed as

$$\mathbf{H} = \begin{bmatrix} h_{11} & \cdots & h_{1N_T} \\ \vdots & \ddots & \vdots \\ h_{N_R1} & \cdots & h_{N_RN_T} \end{bmatrix}. \tag{9}$$

Since the CSI is dynamically changed, the channel matrix $\mathbf{H}$ in (9) is corrected by using the statistical characteristics of the channel. In order to distinguish, we use $\mathbf{H}^D$ to denote the corrected channel. The channel $\mathbf{H}^D$ can be decomposed as into fixed and variable parts as [24]

$$\mathbf{H}^D = \mathbf{H}_m + \overline{\mathbf{H}}, \tag{10}$$

where $\mathbf{H}_m$ is the channel mean and $\overline{H}$ is the changing part of the channel, which consists of the channel instantaneous value and the correlation value. The mean and instantaneous measurements of the channel constitute the estimated value of the channel $\widehat{H}_s$. The correlation value can be represented by the error covariance of the channel $\mathbf{R}_e$. Thus, at the s moment of transmission time, the channel matrix $\mathbf{H}_s$ can be composed of the channel's estimated value $\widehat{H}_s$ and its error covariance $\mathbf{R}_e$, which is expressed as

$$\mathbf{H}_s = \widehat{H}_s + \mathbf{R}_e^{(1/2)}. \tag{11}$$

Assuming that the transmitter has an initial channel measurement $\mathbf{H}_0$ at time 0, together with the channel mean $\mathbf{H}_m$, the channel covariance $\mathbf{R}_0$ which represent the spatial correlation between all pairs of transmit and receive antennas in the MIMO system, and the autocovariance of the channel $\mathbf{R}_s$. Based on minimum mean squared error (MMSE) estimation theory, an optimal estimate $\widehat{h}_s$ of the channel at time s and the estimation error covariance $\mathbf{R}_{e,s}$ can be established as

$$\begin{aligned} \widehat{h}_s &= \mathrm{E}\left[\mathbf{h}_s|\mathbf{h}_0\right] = \mathbf{h}_m + \mathbf{R}_s^* \mathbf{R}_0^{-1}\left[\mathbf{h}_0 - \mathbf{h}_m\right], \\ \mathbf{R}_{e,s} &= \mathrm{cov}\left[\mathbf{h}_s|\mathbf{h}_0\right] = \mathbf{R}_0 - \mathbf{R}_s^* \mathbf{R}_0^{-1}\mathbf{R}_s, \end{aligned} \tag{12}$$

where $\widehat{h}_s = \mathrm{vec}(\widehat{H}_s)$. Based on the premise that the channel temporal statistics can be the same for all antenna pairs, it may be assumed that the temporal correlation is homogeneous and identical for any channel element [25, 26]. The assumption has been used to construct a channel model and verify its existence by measuring the data. Then, the spatial and temporal correlation effects are separable, and the channel autocovariance can be expressed as

$$\mathbf{R}_s = \beta_s \mathbf{R}_0, \tag{13}$$

where $\beta_s = J_0(2\pi f_{\max} T_s)$ is the temporal correlation of channel, which can be obtained according to the Jakes fading model [27, 28]. $J_0(\cdot)$ denotes the first-kind zero-order Bessel function and T_s is the channel sampling time interval. According to (13), we know that all the $NN_R \times N_T$ channels between the N_T transmit and NN_R receive antennas have the same temporal correlation function.

Then, we apply the simplified temporal correlation model (12). This model helps to isolate the effect of temporal channel variation on the CSI. The channel estimate $\widehat{H}_s$ and its error covariance $\mathbf{R}_e$ are rewritten as

$$\begin{aligned} \widehat{H}_s &= \beta_s \mathbf{H}_0 + (1-\beta_s)\mathbf{H}_m, \\ \mathbf{R}_e &= \left(1-\beta_s^2\right)\mathbf{R}_0, \end{aligned} \tag{14}$$

where $\mathbf{H}_m = (1/L)\sum_{j=-L+1}^{0} \mathbf{H}_j$, L is the window length of the CSI mean, and $\mathbf{H}_j$ is the instantaneous channel measurement at j sampling time. Then, the transmitter CSI can be briefly characterized as a function of the temporal correlation coefficient of the channel β_s, the measured value of the channel $\mathbf{H}_0$, the mean of the channel $\mathbf{H}_m$, and covariance of the channel $\mathbf{R}_0$.

$\mathbf{R}_0$ is a semidefinite Hermitian matrix. Its diagonal elements represent the power gain of the $NN_R \times N_T$ channels, and the off-diagonal elements are the cross-coupling between these scalar channels. In the channel model constructed based on the Kronecker structure, $\mathbf{R}_0$ can be decomposed into the Kronecker product of the transmitter antenna correlation $\mathbf{R}_t^T$ and the receiver antenna correlation $\mathbf{R}_r$:

$$\mathbf{R}_0 = \mathbf{R}_t^T \otimes \mathbf{R}_r, \tag{15}$$

where $\mathbf{R}_t$ and $\mathbf{R}_r$ both are semidefinite Hermitian matrix. $\mathbf{R}_t = (1/NN_R)((1/L)\sum_{j=-L+1}^{0} \mathbf{H}_j^* \mathbf{H}_j - \mathbf{H}_m^* \mathbf{H}_m)$ and $\mathbf{R}_r = (1/N_T)((1/L)\sum_{j=-L+1}^{0} \mathbf{H}_j^* \mathbf{H}_j - \mathbf{H}_m^* \mathbf{H}_m)$. Substituting (14) and (15) into (11), the channel matrix $\mathbf{H}^D$ based on dynamic CSI can be obtained:

$$\mathbf{H}^D = \beta_s \mathbf{H}_0 + (1-\beta_s)\mathbf{H}_m + \left[\left(1-\beta_s^2\right)\mathbf{R}_t^T \otimes \mathbf{R}_r\right]^{(1/2)}. \tag{16}$$

3.2. Greedy Algorithm Based on Maximizing Spectrum Utilization. This paper uses greedy algorithms to select and rank users to maximize spectrum utilization while reducing system bit error ratio (BER). Firstly, a user with the largest spectrum utilization is selected from all N users. Then, we select the next user. Each time the newly selected user and the previously selected user are combined to get the largest spectrum utilization. When the number of selected users reaches the upper limit we set, the selection is stopped and the algorithm ends. The specific process of the algorithm is shown in Algorithm 1.

The greedy algorithm used in this section determines whether to select the joint channel composed of each user and the selected user by continuously calculating whether it has the maximum spectrum utilization. Compared with the traversal algorithm, the search numbers are greatly reduced, and the result is equal to or close to the optimal solution. When the greedy algorithm selects a user, the joint channel matrix is generated, which means the user is synchronously sorted. It can improve the system performance to a certain extent.

Input: number of transmitting antennas N_T, signal-to-noise ratio (SNR) P/N_0, number of candidate users N, candidate user channel matrix $\mathbf{H}^D$, number of target users K;
Output: channel matrix of selected users $\mathbf{H}_K$.
Step 1. Initialization: Set candidate user set $\Gamma = \{1, 2, \ldots, N\}$, selected user set $\Omega = \varnothing$, channel matrix $\mathbf{H}_K =$, which is an empty matrix;
Step 2. Calculate the spectrum utilization of each user in the candidate user, and select the user with the largest spectrum utilization. $s_1 = \text{argmax}_{k_1 \in \Gamma} \log_2 \det(\mathbf{I} + (P/N_T N_0)\mathbf{H}_{k1}\mathbf{H}_{k1}^H)$ bps/Hz;
Step 3. Update the sets. $\Gamma = \Gamma - \{s_1\}$, $\Omega = \Omega + \{s_1\}$, and update the channel matrix $\mathbf{H}_K = [\mathbf{H}_{k1}]$;
Step 4. The users in the updated set Γ are combined with the selected users to form a new joint channel matrix. Then, calculate the spectrum utilization of the joint channel, and select the user with the largest combined spectrum utilization;
Step 5. Repeat step 3 until the number of target users is selected;
Step 6. End and output the channel matrix of the selected users $\mathbf{H}_K$.

Algorithm 1: User scheduling based on the greedy algorithm.

Input: system total power P, user number K, the weighting matrix $\mathbf{G}$, the feedforward matrix $\mathbf{F}$, the channel matrix $\mathbf{H}_K$, threshold ε, water filling line μ;
Output: power allocation matrix for each user $\mathbf{P}_K$.
Step 1. Initialize the power distribution, $P_j^{(0)} = (P/K), \quad j = 1, \ldots, K$;
Step 2. Update power $P_j^{(m)} = (\mu_j^m - (\sum_{i=1, i\neq j}^{K} \|g_j \mathbf{h}_j \mathbf{f}_j\|_F^2 P_j + n_{R_j}\|g_j\|_F^2 \sigma^2 / \|g_j \mathbf{h}_j \mathbf{f}_j\|_F^2))^+ j = 1, \ldots, K$, $\sum_{j=1}^{K} P_j^{(m-1)} = P$;
Step 3. If $(|P_j^{(m)} - P_j^{(m-1)}|/P_j^{(m)}) \leq \varepsilon$, $j = 1, \ldots K$, go to step 4, or go back to step 2;
Step 4. Iterated over, output user power allocation matrix $\mathbf{P}_K = \text{diag}(P_1^{(m)}, P_2^{(m)}, \ldots, P_K^{(m)})$;

Algorithm 2: Iterative water filling power allocation algorithm.

3.3. Iterative Water Filling Power Allocation Algorithm. The water filling algorithm is a classic power allocation algorithm [29]. The algorithm allocates power to each user based on the channel characteristics to improve the spectrum utilization of the system.

Figure 2 is a diagram of power allocation for a multiuser MIMO precoding system. $P_j (j = 1, \ldots, K)$ represent the power allocated to each user, $\mathbf{h}_j$ represents the j-th row of the channel matrix $\mathbf{H}_K$, and $\mathbf{f}_j$ represents the j-th column of the feedforward matrix $\mathbf{F}$, so the spectrum utilization can be expressed as

$$C = \sum_{j=1}^{K} \log_2 \left(1 + \frac{\left\| g_j \mathbf{h}_j \mathbf{f}_j \right\|_F^2 P_j}{\sum_{i=1, i\neq j}^{K} \left\| g_j \mathbf{h}_j \mathbf{f}_j \right\|_F^2 P_j + n_{R_j} \left\| g_j \right\|_F^2 \sigma^2} \right). \tag{17}$$

To maximize spectrum utilization, we can construct objective function and constrain as follows:

$$\begin{cases} \max C = \sum_{j=1}^{K} C_j(P_j), \\ \text{s.t.} \sum_{j=1}^{K} P_j = P \text{ and } P_j \geq 0, \quad j = 1, \ldots, K. \end{cases} \tag{18}$$

An iterative water filling algorithm suitable for multiuser MIMO systems is designed. Firstly, the power of each user is initialized, and the power is distributed equally after obtaining the precoding matrix. Then, the user power is iteratively updated according to the objective function. In each iteration, the interference is equivalent to noise and processed, and the user power allocation is optimized through the water filling algorithm. When the power allocated by each user basically does not change, the iteration stops and the algorithm ends. The specific process of the iterative water filling power allocation algorithm is as follows.

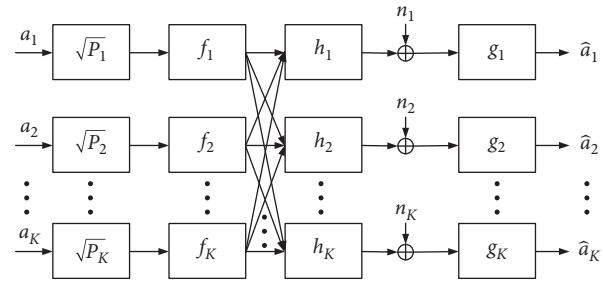


Figure 2: Diagram of power allocation for the multiuser precoding system.

The iterative water filling algorithm takes full account of the influence of interference and treats it as noise. When there is no spatial correlation in the channel, the singular value distribution of each user channel is relatively average, and the system using the iterative water filling algorithm does not have significant channel capacity. If the channel correlation exists, the channel condition number is large, or even the channel rank is not satisfied, the channel capacity advantage of the iterative water filling algorithm will be more obvious.

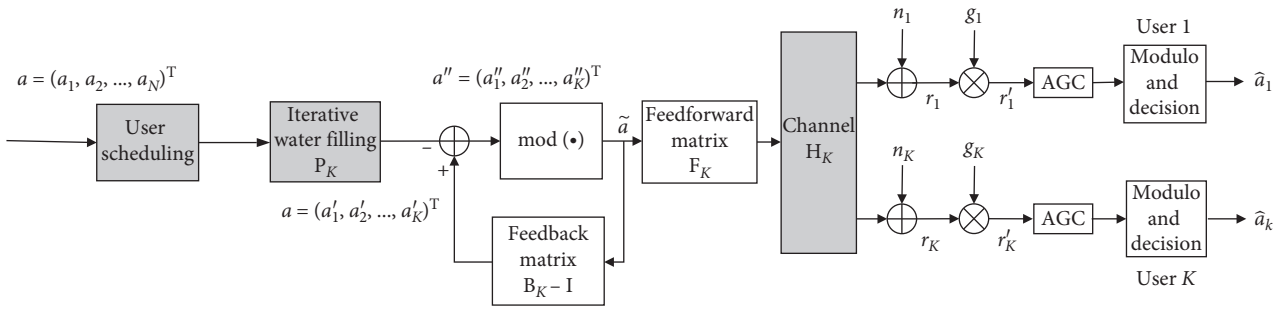


Figure 3: Diagram of improved THP based on the iterative water filling algorithm.

3.4. Precoding Design. The diagram of improved THP based on the iterative water filling algorithm is shown in Figure 3. The original transmission signal is $\mathbf{a} = [a_1, a_2, \ldots, a_N]^T$. The transmission signal after user scheduling and sorting by the greedy algorithm proposed in subsection $\mathscr{B}$ is $a' = [a'_1, a'_2, \ldots, a'_K]^T$, and the transmission signal after power allocation by the iterative water injection algorithm proposed in subsection $\mathscr{C}$ is $a'' = [a''_1, a''_2, \ldots, a''_K]^T$. After the user scheduling, reconstructed channel matrix $\mathbf{H}_K$ will generate the corresponding feedback matrix $\mathbf{B}_K$, a weighting matrix $\mathbf{G}_K$, and a feedforward matrix $\mathbf{F}_K$. Then, the received vector $\mathbf{r}'_K$ is

$$\mathbf{r}'_K = \mathbf{G}_K\left(\mathbf{H}_K\mathbf{F}_K\tilde{a} + \mathbf{n}_K\right) = \mathbf{G}_K\mathbf{H}_K\mathbf{F}_K\mathbf{B}_K^{-1}\mathbf{v}_K + \mathbf{G}_K\mathbf{n}_K, \quad (19)$$

where $\mathbf{r}'_K = [r'_1, r'_2, \ldots, r'_K]^T$, $\mathbf{n}'_K = [n_1, n_2, \ldots, n_K]^T$, and $\mathbf{v}'_K = \mathbf{B}'_K\tilde{a}$ are the equivalent transmit data vector. As can be seen from Figure 3, the error $\mathbf{e}$ between the received vector $\mathbf{r}'_K$ and the equivalent transmitted vector $\mathbf{v}_K$ is expressed as

$$\begin{aligned}\mathbf{e} &= \mathbf{r}'_K - \mathbf{v}'_K \\ &= \mathbf{G}_K\left(\mathbf{H}_K\mathbf{F}_K\tilde{a} + \mathbf{n}_K\right) - \mathbf{v}_K \\ &= \left(\mathbf{G}_K\mathbf{H}_K\mathbf{F}_K - \mathbf{B}_K\right)\tilde{a} + \mathbf{G}_K\mathbf{n}_K.\end{aligned} \quad (20)$$

According to the MMSE algorithm, we can construct objective functions and constraints as follows:

$$\begin{cases} \underset{\mathbf{B}_K,\mathbf{G}_K,\mathbf{F}_K}{\arg\min}\, E\left[\left\|\left(\mathbf{G}_K\mathbf{H}_K\mathbf{F}_K - \mathbf{B}_K\right)\tilde{a} + \mathbf{G}_K n_K\right\|^2\right], \\ \text{s.t. } \|\tilde{a}\|^2 \le P, \end{cases} \quad (21)$$

where P represents the total transmit power at the transmitter. By using the orthogonal principle, we have

$$\begin{aligned}E\left[\mathbf{e}\mathbf{r}_K^H\right] &= E\left[\left(\mathbf{r}'_K - \mathbf{v}_K\right)\mathbf{r}_K^H\right] \\ &= E\left[\mathbf{G}_K\mathbf{r}_K\mathbf{r}_K^H - \mathbf{B}_K\tilde{a}\mathbf{r}_K^H\right] \\ &= 0.\end{aligned} \quad (22)$$

Substituting (20) into (22), we have

$$\mathbf{G}_K\left(\mathbf{H}_K\mathbf{F}_K\mathbf{F}_K^H\mathbf{H}_K^H + \xi\mathbf{I}_K\right) = \mathbf{B}_K\mathbf{F}_K^H\mathbf{H}_K^H, \quad (23)$$

where $\xi = (\sigma^2_{\mathbf{n}_K}/\sigma^2_{\tilde{a}}) = (E[\mathbf{n}_K\mathbf{n}_K^H]/E[\tilde{a}\tilde{a}^H])$. Assuming that the feedforward matrix $\mathbf{F}_K$ is a unitary matrix which satisfies $\mathbf{F}_K\mathbf{F}_K^H = \mathbf{I}$, then we have

$$\mathbf{G}_K^{-1}\mathbf{B}_K\mathbf{B}_K^H\mathbf{G}_K^{-H} = \left(\mathbf{H}_K\mathbf{H}_K^H + \xi\mathbf{I}\right)\mathbf{H}_K^{-H}\mathbf{H}_K^{-1}\left(\mathbf{H}_K\mathbf{H}_K^H + \xi\mathbf{I}\right)^H. \quad (24)$$

Let $\mathbf{L}_K = \mathbf{G}_K^{-1}\mathbf{B}_K$, then (25) can be expressed as

$$\mathbf{L}_K\mathbf{L}_K^H = \left(\mathbf{H}_K\mathbf{H}_K^H + \xi\mathbf{I}\right)\mathbf{H}_K^{-H}\mathbf{H}_K^{-1}\left(\mathbf{H}_K\mathbf{H}_K^H + \xi\mathbf{I}\right)^H. \quad (25)$$

We perform a LQ decomposition of $(\mathbf{H}_K\mathbf{H}_K^H + \xi\mathbf{I})\mathbf{H}_K^{-H}$ and obtain the main diagonal element of the lower triangular matrix $\mathbf{L}_K$. Then, we can obtain the precoding matrices.

$$\begin{aligned}\mathbf{G}_K &= \begin{bmatrix} l_{11}^{-1} & & & \\ & l_{22}^{-1} & & \\ & & \ddots & \\ & & & l_{KK}^{-1} \end{bmatrix}, \\ \mathbf{B}_K &= \mathbf{G}_K\mathbf{L}_K, \\ \mathbf{F}_K &= \mathbf{Q}_K.\end{aligned} \quad (26)$$

4. Complexity and Simulation Analysis

This section mainly analyzes the complexity of the proposed algorithm and several other classical precoding algorithms. The simulation is performed by Matlab, and finally the simulation results are analyzed.

4.1. Complexity Analysis. We compare the computational complexity of the classical linear precoding algorithm, the THP, and the improved THP algorithm in Table 1. It should be noted that the complexity in this paper mainly refers to the number of times the algorithm performs the product operation once.

The zero force (ZF) and MMSE algorithms in Table 1 are classical linear precoding algorithms. The linear precoding algorithm has lower complexity, but its BER performance is poor and the spectrum utilization is low. Compared with ZF, MMSE has a matrix inverse operation, so its complexity is slightly higher than the ZF. The traditional THP algorithm without user scheduling introduces QR decomposition of matrix, so its complexity is higher than that of linear algorithms. Furthermore, the user-scheduled THP algorithm introduces a user scheduling algorithm based on the greedy algorithms, which is less complex than the traversal search. The THP algorithm with the dynamic CSI model introduces

Table 1: Complexity comparison.

Algorithm	Complexity
ZF [30]	N_T^3
MMSE [30]	$4N_T^3 + N_T$
ZF-THP (without user scheduling) [16]	$3N_T^3 + 3N_T^2 + N_T$
MMSE-THP (without user scheduling) [16]	$7N_T^3 + 3N_T^2 + 2N_T$
ZF-THP (with user scheduling)	$3N_T^3 + 3N_T^2 + N_T + K(K+1)(2K+1)(N_T/6)$
MMSE-THP (with user scheduling)	$7N_T^3 + 3N_T^2 + 2N_T + K(K+1)(2K+1)(N_T/6)$
ZF-THP (dynamic CSI model)	$N_T^4 + 3N_T^3 + 3N_T^2 + 3N_T$
MMSE-THP (dynamic CSI model)	$N_T^4 + 7N_T^3 + 3N_T^2 + 4N_T$
The proposed algorithm (ZF)	$N_T^4 + 3N_T^3 + 3N_T^2 + 3N_T + K(K+1)(2K+1)(N_T/6)$
The proposed algorithm (MMSE)	$N_T^4 + 7N_T^3 + 3N_T^2 + 4N_T + K(K+1)(2K+1)(N_T/6)$

Kronecker product operation with complexity N_T^4, so its complexity is higher than traditional THP algorithm. The proposed algorithm not only adds the dynamic CSI model and user scheduling but also adds the iterative water filling power allocation algorithm, so the complexity is higher than other algorithms. However, from a practical point of view, the computational complexity of the proposed algorithm is still within acceptable limits.

Table 2: Simulation parameters.

Parameters	Value
Carrier frequency	5.9 GHz
Moving speed	30 km/h, 120 km/h
Transmitting antennas N_T	4
Receiving antennas N_R	1
Total users N	10
Target users K	4
Modulation	QPSK

4.2. Simulation Analysis. In order to verify the rationality of the proposed algorithm, we construct a MIMO-V2I communication system to simulate and verify the proposed algorithm in a high-speed mobile scenario. In the simulation, a 4×1 multiuser MIMO model is adopted. The channel model is WINNER II channel [23]. The simulation system parameters are shown in Table 2.

This paper mainly introduces BER and spectrum utilization of linear precoding [30], the traditional THP algorithm [16], and the proposed algorithm in different speed scenarios.

Figures 4(a) and 4(b) show the BER of linear precoding, traditional THP, and the proposed algorithm when the vehicle speed is 120 km/h and 30 km/h.

In high-speed scenarios, the proposed algorithm is the best. The traditional THP algorithm based on the dynamic CSI model is second, and the traditional THP algorithm combined with the user scheduling algorithm is better than the traditional THP. The linear precoding is the worst. This is because the dynamic CSI model compensates for the instantaneous channel through channel correlation, so the CSI model obtained through feedback is closer to the actual channel, thereby improving system performance. The user scheduling algorithm based on the greedy algorithm preferentially selects the user with the highest spectrum utilization rate for priority transmission, which not only improves the spectrum utilization rate of the system but also effectively filters out noise and interference, thereby reducing BER.

In low-speed scenarios, the BER of various precoding algorithms is consistent with the trend of BER in high-speed scenarios. We observe that the precoding algorithm combined with the dynamic CSI model is close to that of the precoding algorithm without the dynamic CSI model. This is because the moving speed is low, there is no strong correlation between channels, and the BER advantage brought by the dynamic CSI model is not obvious.

In summary, the proposed algorithm is suitable for high-speed mobile scenarios.

Figures 5(a) and 5(b) show the spectrum utilization of linear precoding, traditional THP, and the proposed algorithm at a vehicle speed of 120 km/h and 30 km/h.

In high-speed scenarios, the proposed algorithm has the best spectrum utilization performance. This is because the multiuser scheduling algorithm based on the greedy algorithm optimizes the target by maximizing the channel capacity, thereby improving the spectrum utilization of the system. Moreover, the dynamic CSI model makes the channel closer to the actual channel, and the accurate CSI can design a better precoding matrix and further improve the spectrum utilization.

In low-speed scenarios, the trend of precoding algorithm spectrum utilization is consistent with that in high-speed scenarios. However, in low-speed scenarios, the spectrum utilization of each precoding algorithm is higher than that of similar algorithms in high-speed situations. According to information theory, when the SNR is determined, the spectrum utilization is only related to the channel matrix. In the low-speed scene, the channel is corrected by channel correlation, and the channel gain is improved so that the spectrum utilization is effectively improved to a certain extent.

Figure 6 shows the spectrum utilization comparison between the precoding algorithm using the iterative water filling algorithm and the precoding algorithm using equal power allocation. As illustrated, spectrum utilization is improved by using iterative water filling power allocation. The adopted iterative water filling algorithm carries out adaptive power allocation for users. More power is allocated

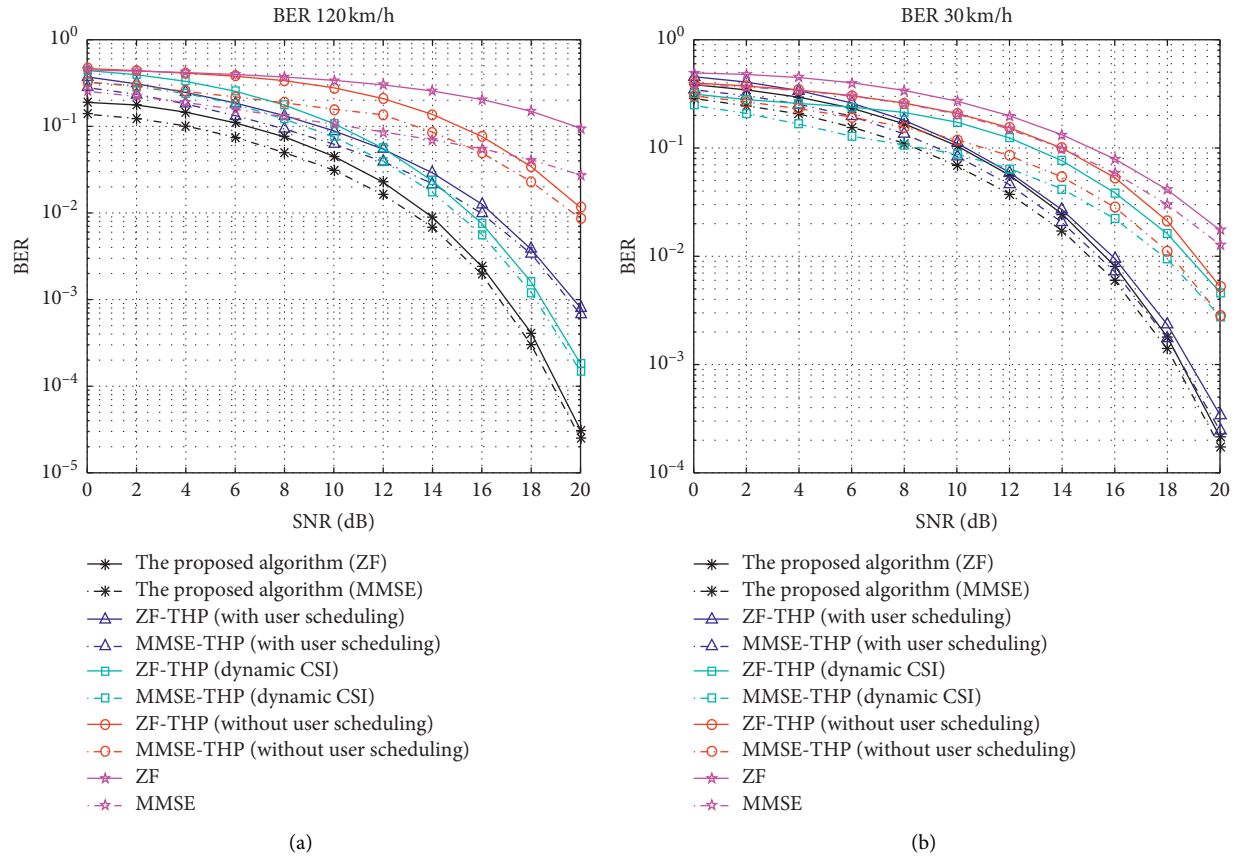


Figure 4: BER of several precoding algorithms: (a) 120 km/h; (b) 30 km/h.

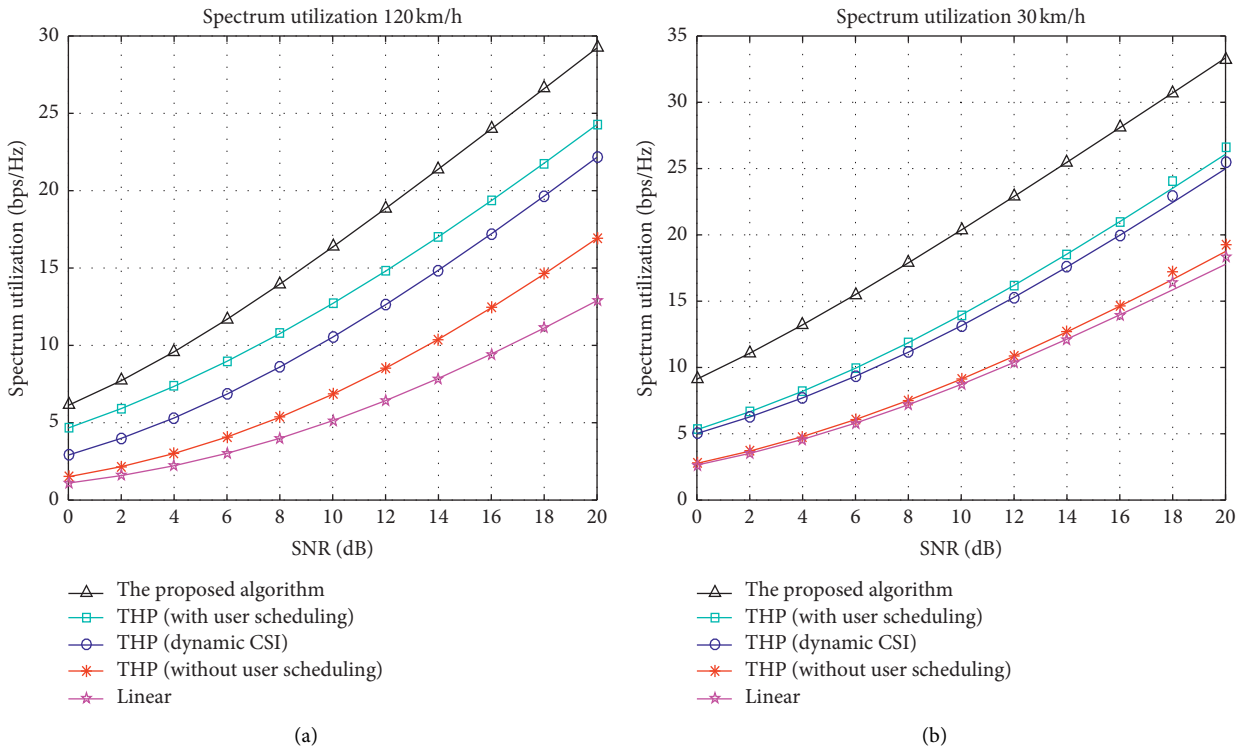


Figure 5: Spectrum utilization of several precoding algorithms: (a) 120 km/h; (b) 30 km/h.

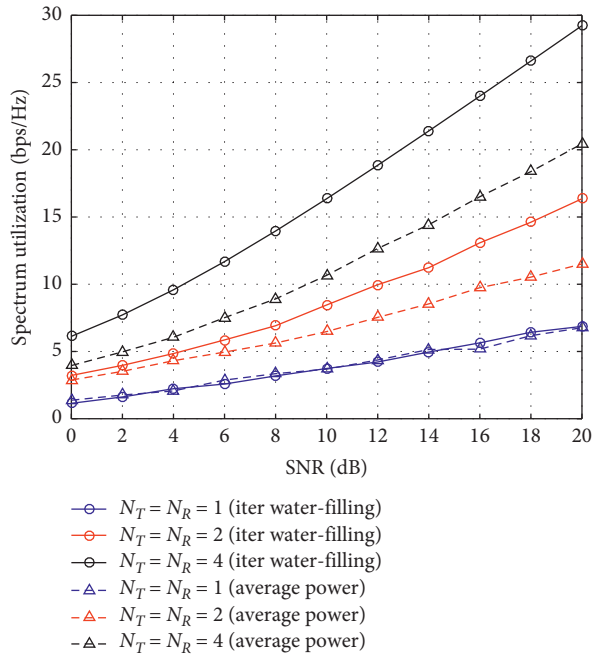


Figure 6: Spectrum utilization under different power allocation algorithms.

to users with good channel conditions while less or no power is allocated with poor channel conditions. The power resources are effectively utilized so that the spectrum utilization of the system is improved as a whole. At the same time, with the increase in the antennas number, the spectrum utilization of the iterative water filling power allocation algorithm has obvious advantages. It is verified that the iterative water filling power algorithm is suitable for the MIMO systems with multiple antennas at the transmitter and receiver.

5. Conclusion

This paper studies the precoding of the ultra reliable communication for V2I of CPSs. Aiming at the fast time-varying characteristics of the channel in high-speed mobility scenarios, we propose an improved THP algorithm. By constructing a dynamic CSI model based on channel statistic information and related characteristics, the obtained dynamic CSI is closer to the current actual channel. At the same time, the iterative water filling power allocation algorithm and the user scheduling algorithm based on the greedy algorithm are combined with the THP algorithm to carry out joint optimization design. The simulation results show that the proposed algorithm has better BER and spectrum utilization than the traditional THP algorithm and linear precoding algorithm and is more suitable for fast time-varying channel environment in high-speed mobility scenarios.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 61501066) and Natural Science Foundation of Chongqing (no. cstc2019jcyj-msxmX0017).

References

[1] Y. Zhou, F. R. Yu, J. Chen, and Y. Kuo, "Cyber-physical-social systems: a state-of-the-art survey, challenges and opportunities," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 389–425, 2020.

[2] Q. Sun, Y. Zhang, H. He, D. Ma, and H. Zhang, "A novel energy function-based stability evaluation and nonlinear control approach for energy internet," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1195–1210, 2017.

[3] Q. Sun, S. Chen, L. Chen, and D. Ma, "Quasi-Z-source network-based hybrid power supply system for aluminum electrolysis industry," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1141–1151, 2017.

[4] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 383–398, 2019.

[5] R. Ghosh, R. Pragathi, S. Ullas, and S. Borra, "Intelligent transportation systems: a survey," in *Proceedings of the 2017 International Conference on Circuits, Controls, and Communications (CCUBE)*, pp. 160–165, Bangalore, India, December 2017.

[6] W. Hu, C. Ruan, H. Nian, and D. Sun, "Zero-sequence current suppression strategy with common-mode voltage control for open-end winding PMSM drives with common DC bus," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 6, pp. 4691–4702, 2021.

[7] W. Rui, S. Qiuye, M. Dazhong, and H. Xuguang, "Line impedance cooperative stability region identification method for grid-tied inverters under weak grids," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 2856–2866, 2020.

[8] R. Wang, Q. Sun, W. Hu, Y. Li, D. Ma, and P. Wang, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, pp. 7866–7879, 2021.

[9] R. Wang, Q. Sun, D. Ma, and Z. Liu, "The small-signal stability analysis of the droop-controlled converter in electromagnetic timescale," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1459–1469, 2019.

[10] T. Wang, X. Wang, Z. Cui, Y. Cao, and C. Suthaputchakun, "Survey on cooperatively V2X downloading for intelligent transport systems," *IET Intelligent Transport Systems*, vol. 13, no. 1, pp. 13–21, 2019.

[11] I. Shaer, A. Haque, and A. Shami, "Multi-component V2X applications placement in edge computing environment," in *Proceedings of the 2020 IEEE International Conference on Communications (ICC 2020)*, pp. 1–6, Dublin, Ireland, June 2020.

[12] S. Moon and I. Hwang, "Design and performance analysis of vehicle MIMO system for NR-based enhanced V2X," in *Proceedings of the 2018 International Conference on*

Information and Communication Technology Convergence (ICTC), pp. 351–355, Jeju, Korea (South), October 2018.

[13] B. Wang, R. Shi, C. Ji, and J. Hu, "Joint precoding and user scheduling for full-duplex cooperative MIMO-NOMA V2X networks," in *Proceedings of the IEEE Vehicular Technology Conference (VTC2019-Fall)*, pp. 1–6, Honolulu, HI, USA, September 2019.

[14] Y. Wu, C. Xiao, Z. Ding, X. Gao, and S. Jin, "A survey on MIMO transmission with finite input signals: technical challenges, advances, and future trends," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1779–1833, 2018.

[15] E. Castañeda, A. Silva, A. Gameiro, and M. Kountouris, "An overview on resource allocation techniques for multi-user MIMO systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 239–284, 2017.

[16] X. Guo, D. Yang, H. Wang, J. Kuang, and X. Wen, "MMSE-THP with QoS requirements for the downlink of multiuser MIMO systems," in *Proceedings of the 2018 88th IEEE Vehicular Technology Conference (VTC-Fall)*, pp. 1–5, Chicago, IL, USA, August 2018.

[17] M. Vu and A. Paulraj, "On the capacity of MIMO wireless channels with dynamic CSIT," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1269–1283, 2007.

[18] J. Wang, M. Bengtsson, B. Ottersten, and A. P. Palomar, "Robust MIMO precoding for several classes of channel uncertainty," *IEEE Transactions on Signal Processing*, vol. 61, no. 12, pp. 3056–3070, 2013.

[19] C. Kotchasarn, "Power allocation for multi-user downlink MIMO transmissions," in *Proceedings of the IEEE International Conference on Innovative Research and Development (ICIRD)*, pp. 1–5, Bangkok, Thailand, May 2018.

[20] K. A. Bonsu, W. Zhou, S. Pan, and Y. Yan, "Optimal power allocation with limited feedback of channel state information in multi-user MIMO systems," *China Communications*, vol. 17, no. 2, pp. 163–175, 2020.

[21] R. Zhang, B. Ai, L. Yang, H. Song, and Z.-Q. Li, "A precoding and detection scheme for OFDM based wireless communication system in high-speed environment," *IEEE Transactions on Consumer Electronics*, vol. 60, no. 4, pp. 558–566, 2014.

[22] M. Benmimoune and D. Massicotte, "Multi-user MIMO precoding with Kerdock codebook," in *Proceedings of the 2010 7th International Symposium on Wireless Communication System*, pp. 71–75, York, UK, September 2010.

[23] A. Ghazal, Y. Yuan, C.-X. Wang et al., "A non-stationary IMT-advanced MIMO Channel Model for high-mobility wireless communication systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2057–2068, 2017.

[24] M. Vu and A. Paulraj, "Optimal linear precoders for MIMO wireless correlated channels with nonzero mean in space-time coded systems," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2318–2332, 2006.

[25] Y. Li, X. Cheng, and N. Zhang, "Deterministic and stochastic simulators for non-isotropic V2V-MIMO wideband channels," *China Communications*, vol. 15, no. 7, pp. 18–29, 2018.

[26] Y. Yuan, C.-X. Wang, Y. He, M. M. Alwakeel, and E.-H. M. Aggoune, "3D wideband non-stationary geometry-based stochastic models for non-isotropic MIMO vehicle-to-vehicle channels," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6883–6895, 2015.

[27] C. Kong, C. Zhong, A. K. Papazafeiropoulos, M. Matthaiou, and Z. Zhang, "Sum-rate and power scaling of massive MIMO systems with channel aging," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4879–4893, 2015.

[28] L. Zhao, K. Zheng, H. Long, H. Zhao, and W. Wang, "Performance analysis for downlink massive multiple-input multiple-output system with channel state information delay under maximum ratio transmission precoding," *IET Communications*, vol. 8, no. 3, pp. 390–398, 2014.

[29] P. He, L. Zhao, and B. Venkatesh, "Novel water-filling for maximum throughput of power grid, MIMO, and energy harvesting coexisting system with mixed constraints," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 827–838, 2017.

[30] T. K. Lyu, "Capacity of multi-user MIMO systems with MMSE and ZF precoding," in *Proceedings of the 2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 1083-1084, San Francisco, CA, USA, April 2016.

Hindawi
Complexity
Volume 2021, Article ID 5539829, 16 pages
https://doi.org/10.1155/2021/5539829

WILEY | Hindawi

Research Article

An Optimal DoS Attack Strategy Disturbing the Distributed Economic Dispatch of Microgrid

Yihe Wang,[1] **Mingli Zhang**,[2] **Kun Song**,[2] **Tie Li**,[3] **and Na Zhang**[2]

[1]*School of Information Science and Engineering, Northeastern University, Shenyang 110819, China*
[2]*State Grid Liaoning Electric Power Company Limited Economic Research Institute, Shenyang 110015, China*
[3]*State Grid Liaoning Electric Power Supply Co. Ltd., Shenyang 110004, China*

Correspondence should be addressed to Yihe Wang; 1710247@stu.neu.edu.cn

Received 20 January 2021; Revised 15 February 2021; Accepted 6 March 2021; Published 2 April 2021

Academic Editor: Rui Wang

As a promising method with excellent characteristics in terms of resilience and dependability, distributed methods are gradually used in the field of energy management of microgrid. However, these methods have more stringent requirements on the working conditions, which will make the system more sensitive to communication failures and cyberattacks. As a result, it is both theoretical merits and practical values to investigate the malicious effect of cyber attacks on microgrid. This paper studies the distributed economic dispatch problem (EDP) under denial-of-service (DoS) attacks for the microgrid, in which each generator can communicate with its neighbors and has the computational capability to implement local operation. Firstly, a DoS attack model is proposed, in which the DoS attacker intentionally jams the communication channel to deteriorate the performance of the microgrid. Then, the evolution mechanism of the dispatch system of the microgrid under different attack scenarios is adequately discussed. On this basis, an optimal attack strategy based on enumerating-search algorithm is presented to allocate the limited attack resources reasonably, so as to maximize the effect of DoS attacks. Finally, the validity of the theoretical studies about the attack effect under different scenarios and the effectiveness of the proposed enumerating-search-based optimal attack strategy are illustrated through the simulation examples on the IEEE 57-bus system and IEEE 39-bus system, respectively.

1. Introduction

The intelligent microgrid integrates various renewable generators and energy storage systems within a distributed system to achieve the goal of energy-saving and efficient operation and has been widely recognized as the future of power system. As a multidimensional complex system, the microgrid includes widely distributed sensors, capacity of control and optimization computation, and well-developed communication networks [1]. Now, it has been widely recognized as a typical cyber-physical system (CPS), and various fundamental problems of the microgrid have been reinvestigated in the perspective of CPS in the last few years, such as droop control [2], economic dispatch [3], and zero-sequence current suppression control [4].

Like the traditional power system, economic dispatch is one of the most concerned issues in the operation of the microgrid. Economic dispatch of the microgrid with the goal of minimizing operating costs can realize the collaborative optimization and energy distribution among units. Generally speaking, the methods to solve the economic dispatch problem (EDP) of the microgrid can be classified into centralized methods and distributed methods. Although the centralized methods are efficient, they need central authorities to collect global information of all the controlled units. These methods are sensitive to single-point failure and increase the communication burden of system [5]. On the opposite, the distributed methods without central authorities can carry out the economic dispatch with less communication burden and better scalability and privacy [6]. As a consequence, scholars have investigated the EDP for the microgrid based on various distributed methods in the past few years [7–10]. For example, a fully distributed algorithm proposed in [7] can fit well with arbitrary initializations. Moreover, it enabled generators to collectively obtain the mismatch between demand and total power output as a feedback mechanism to adjust actual power

generation. In order to maximize the social welfare, a distributed consensus algorithm based on augmented Lagrangian and projection penalty function was proposed in [8]. Literature [9] presented a distributed double-Newton descent algorithm for cooperative energy management with faster convergence speed in a fully distributed fashion. In [10], a novel distributed primal-dual augmented subgradient algorithm was investigated, which can solve a class of distributed constrained optimization problems over a general unbalanced directed communication network. However, the above literature assumes that these algorithms are executed under ideal communication conditions and the possible cyberattacks in the communication network are not considered. In practice, once the malicious cyberattacks occur in the communication network of the microgrid, the data transmission will be disturbed, and the execution of those algorithms will be affected. Consequently, the optimal operation points of the system cannot be obtained, and even the stability is destroyed [11]. Therefore, it is of theoretical and practical significance to investigate the EDP under various cyberattacks.

Until now, two types of cyberattacks have received the most attention: denial-of-service (DoS) attacks and deception attacks [12]. As a common type of cyberattacks in CPS, DoS attacks attempt to make the network resource unavailable for its intended users. This kind of attack blocks real traffic by sending a mass of invalid information that takes up all the communication bandwidth. Currently, research on DoS attacks has two main perspectives as follows. (1) From the perspective of system defense, in order to mitigate the vulnerability of system under attacks, many robust methods have been investigated in recent years [13–15], such as event/self-triggered control [13], the confidence-level-based resilient strategy [14], and intensity-dependent consensus protocol [15]. (2) From the perspective of DoS attacker, in order to increase the probability of successful attacks and enhance the effect of attacks, the attacker tends to formulate various optimal attack strategies according to their working conditions. Recently, there are some studies on formulating optimal attack strategies. For example, literature [16] proposed an optimal attack strategy under limited attack resources, and this strategy can damage the estimation performance of the system by deciding which sensor to attack. And, on this basis, the optimal attack time and attack mode are determined in [17] to maximum the impact of DoS attacks on the remote state estimation of the power system. However, the above research on the optimal attack strategies only discusses the DoS attacks against the state estimation, and the goal of it is to undermine the security and stability of systems. Until now, there are few literatures on designing the optimal DoS attacks strategies to damage the economic performance of the power system. However, the degradation caused by DoS attacks to the economic performance of the power system cannot be ignored, which may have a great impact on the national economy. As a result, it is necessary to study under what strategy the attack will cause the maximum degree of economic performance degradation, i.e., the optimal attack strategy disturbing the distributed economic dispatch of the microgrid.

Based on the above analysis, the motivation of the current work is in the following three aspects:

(1) As a common type of cyberattacks, DoS attacks are highly destructive, which may cause severe consequences for the power system.

(2) As far as we know, there are few literatures about the impact of DoS attacks on the economic dispatch of the power system. However, it is essential to analyze potential consequences of these attacks on the economic performance of the system.

(3) Research on optimal attack strategies can guide the formulation of defense measures for energy systems, and one needs to understand what the worst effect of attacks might be.

In this paper, we investigate the economic dispatch problem of the microgrid under DoS attacks. Firstly, the model of DoS attacks is established according to their working mechanism. Then, the impact of DoS attacks on the economic dispatch of the microgrid under different scenarios is investigated. On this basis, an optimal attack strategy that can maximize the effect of attacks with limited attack resources is proposed. The main contributions of this paper are summarized as follows:

(1) In this paper, the EDP of the microgrid under DoS attacks is investigated, which further enriches the theoretical research on DoS attacks in the power system.

(2) The time-varying communication network caused by DoS attacks can be divided into two types, i.e., connected and disconnected. How the two scenarios of attacks affect economic dispatch is analyzed in this paper (Theorems 1 and 2).

(3) Different from the existing literatures, where optimal attack strategies were always put forward to increase the state estimation error of the power system, the optimal attack strategy proposed in this paper aims at damaging economic performance of the microgrid.

(4) An enumerating-search algorithm is proposed to find out the optimal attack strategy for DoS attacker (Algorithm 1).

The rest of the paper is organized as follows. Section 2 revisits some preliminaries, and the problems of interest are summarized. Section 3 illustrates the degradation of system economic performance under the different scenarios of attacks through complete theoretical analysis. Section 4 derives the optimization model of attack strategy and its corresponding algorithm to find out the solution. The simulation examples and conclusion are appeared in Sections 5 and 6, respectively.

2. Preliminaries

2.1. Graph Theory. A graph G is always used to model the network topology of a system. A graph can be denoted as

Input: $P_D; \alpha_i; \beta_i; P_i^{\max}; P_i^{\min}; P_{cut} = 0; \delta_{\max} = 0; \phi; \Lambda$
Compute $\lambda^* = (P_D + \sum_{i=1}^{N} (\beta_i/2\alpha_i)/\sum_{i=1}^{N} (1/2\alpha_i))$
Enumerate the set of all attack schemes $X = \{X_1, X_2, \ldots\}$ according to constraints (31) and (33), and each attack strategy X_j can be further represented as $X_j = \{i|\exists i \in V\}$, which represents the units' isolated set caused by an attack action
Define $m = \text{card}|X|$
  for $j = 1{:}m$ do
    for $i = 1{:}N$ do
      if $i \in X_j$ then
        Compute total output lower limit of isolated generators $P_{\text{cut}} = P_{\text{cut}} + P_i^{\min}$;
      else
        Compute $(\beta/2\alpha) = (\beta/2\alpha) + (\beta_i/2\alpha_i)$;
        Compute $(1/2\alpha) = (1/2\alpha) + (1/2\alpha_i)$;
      end if
    end for
    Compute $\delta = \left| (P_D - P_{\text{cut}} + (\beta/2\alpha)/(1/2\alpha)) - \lambda^* \right|$;
    if $\delta > \delta^{\max}$ then
      Update $\delta^{\max} = \delta$;
      Update X^* according to X_j and ϕ;
    end if
  end for
Output: $\delta^{\max}$; X^*.

Algorithm 1: Solution of optimal attack strategy.

$G = \{V, E, A\}$, where $V = \{1, 2, 3 \ldots\}$ is a finite set of points, and $E \subseteq V \times V$ is an edge set of pairs of nodes. In a directed graph, the edge from i to j is denoted by an ordered pairs $(i, j) \in E$, which means information can be transferred from i to j. $A = [a_{ij}]$ is a matrix named adjacency matrix, a_{ij} denotes the weight of the edge (i, j). If the agent i and j can communicate with each other, they will be defined as neighbors. The in-neighbors of the i^{th} node are denoted by the set $N_i^+ = \{j|(j, i) \in E\}$. Similarly, the out-neighbors of the i^{th} node are denoted by the set $N_i^- = \{j|(i, j) \in E\}$. Especially, each vertex belongs to both its in-neighbor and out-neighbor, i.e., $i \in N_i^+$ as well as $i \in N_i^-$. The in-degree and out-degree of the i^{th} vertex are defined as $d_i^+ = |N_i^+|$ and $d_i^- = |N_i^-|$, respectively. Strongly connected graph can be defined if there are paths between any pair of two vertices in a directed graph, i.e., each vertex is reachable for all the other vertices.

2.2. Stochastic Matrix. A nonnegative square matrix M is defined as row stochastic matrix if $M\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is a column vector with all the elements equal to 1. Similarly, a nonnegative square matrix N is defined as column stochastic matrix if $\mathbf{1}^T N = \mathbf{1}^T$, and a nonnegative square matrix H is double stochastic matrix if it satisfies both $H\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T H = \mathbf{1}^T$.

Moreover, based on the Perron–Frobenius theorem [18], for a row stochastic matrix M, there exists such an identity column vector μ satisfying $\mu^T M = \mu^T$. Similarly, for a column stochastic matrix N, there also exists an identity column vector ξ satisfying $N\xi = \xi$.

2.3. Analytic Solution to EDP. The microgrid is essentially a typical multiagent system, in which each generator can be regarded as an agent. The operation cost function of the i^{th} generation is given by the following quadratic form [19]:

$$C_i(P_i) = \alpha_i P_i^2 + \beta_i P_i + \gamma_i, \tag{1}$$

where P_i is the power generated by generator i and β_i, γ_i, and $\alpha_i > 0$ are the operation cost parameters of generator i.

Assume there is a microgrid system with n generators, the objective of the economic dispatch of the microgrid is to minimize the total operation cost, which is given by

$$\min \sum_{i=1}^{n} C_i(P_i). \tag{2}$$

The EDP of the microgrid needs to be solved under the power balance constraint and power generation constraint as follows:

$$\sum_{i=1}^{n} P_i = P_D, \tag{3}$$

$$P_i^{\min} \le P_i \le P_i^{\max}, \quad \text{for } i = 1, 2, 3 \ldots n, \tag{4}$$

where $P_i^{\min}$ and $P_i^{\max}$ are the lower bound and upper bound of the output power associated with the ith generation, respectively. P_D denotes the total power demand satisfying $\sum_{i=1}^{n} P_i^{\min} \le P_D \le \sum_{i=1}^{n} P_i^{\max}$, i.e., the problem is solvable.

The above optimization problem can be solved by the Lagrange multiplier method [20]. The Lagrange multiplier can be expressed as follows:

$$\lambda_i = \frac{\text{tial}C_i(P_i)}{\text{tial}P_i} = 2\alpha_i P_i + \beta_i. \tag{5}$$

Because the Lagrange multiplier λ_i is equal to the incremental cost $(\text{tial}C_i(P_i)/\text{tial}P_i)$, we directly use λ_i to

represent the incremental cost of units in the following paragraphs.

With the consideration of generation constraint in (4), the power generated by generator i is given by

$$\begin{cases} 2\alpha_i P_i + \beta_i = \lambda^*, & \text{for } P_i^{\min} \le P_i \le P_i^{\max}, \\ 2\alpha_i P_i + \beta_i < \lambda^*, & \text{for } P_i = P_i^{\max}, \\ 2\alpha_i P_i + \beta_i > \lambda^*, & \text{for } P_i = P_i^{\min}, \end{cases} \tag{6}$$

where λ^* is the incremental cost corresponding to the optimal solution of the EDP.

2.4. Problems of Interest. Based on a well-developed distributed algorithm to solve the EDP of the smart grid, the main problems we are interested in consist of the following:

(1) What is the possible effect of DoS attacks on EDP when the attacker can be detected hardly by the monitor in the smart grid?

(2) How to quantify the degradation of economic performance caused by DoS attacks?

(3) Does there exists an optimal attack strategy that renders maximum degradation of economic performance caused by the attacker?

3. Performance Analysis of Dispatch System under DoS Attacks

3.1. DoS Attacks' Model. Assume the DoS attacker is a single source of attacks in the communication network environment of the microgrid, and the DoS attacker sends the invalid requests to occupy the network resources. The DoS attacks target various components including smart meters and communication links, resulting in limited availability of these components. In this paper, the objectives of DoS attacks are to hinder the transmission of information on the communication links, thereby degrading the economic performance of the microgrid.

Stochastic DoS attacks can occur at any communication moment and any communication link. To clearly describe the effect of DoS attacks on the economic dispatch of the microgrid, the model of DoS attacks needs to be established. Define an edge set $U^e(k) \subseteq E$ to represent the communication links attacked at iteration k, $k \in \{0, 1, \ldots\}$. The information transmission over the compromised communication links is impeded.

The effect of DoS attacks on the communication network is equal to packet dropout, which fails the information transmission among agents. Actually, the communication path $(i, j) \in E$ can be regarded as unconnected if the information transmission on the edge is failed. It is reasonable to assume that the network topology could be time-varying and possibly disconnected when the communication network is attacked by the DoS attacker. In this paper, EDP over unreliable network with DoS attacks is regarded as the EDP under time-varying communication topology. Moreover, we assume that once DoS attacks disappear, any disconnected links caused by attacks will be recovered into the connected state by communication restoration mechanism.

$A(k) = [a_{ij}(k)] \in \mathbb{N}^{n\times n}$ is the adjacency matrix corresponding to the communication network topology under DoS attacks at iteration k. It can be defined as the following form:

$$a_{ij}(k+1) = \begin{cases} 0, & (i, j) \in U^e(k+1), \\ 1, & (i, j) \notin U^e(k+1). \end{cases} \tag{7}$$

The updating matrices $M(k) = [m_{ij}(k)], N(k) = [n_{ij}(k)] \in \mathbb{R}^{n\times n}$ can be defined according to $A(k)$ as follows:

$$m_{ij}(k) = \begin{cases} \dfrac{1}{d_i^+(k)}, & j \in N_i^+, \\ 0, & \text{otherwise} \end{cases} \quad \forall i, j \in V,$$
$$n_{ij}(k) = \begin{cases} \dfrac{1}{d_j^-(k)}, & i \in N_j^-, \\ 0, & \text{otherwise,} \end{cases} \quad \forall i, j \in V. \tag{8}$$

It can be deduced that $M(k)$ is a row stochastic matrix and $N(k)$ is a column stochastic matrix.

Lemma 1. *Based on the definition, row stochastic matrix M and column stochastic matrix N both have the eigenvalues 1, and the rest eigenvalues are within the unit circle on the complex plane.*

In this paper, a fully distributed algorithm proposed in [20] to solve the EDP of power system can be transformed into the following form with the consideration of DoS attacks:

$$\lambda_i(k+1) = \sum_{j=1}^{n} m_{ij}(k)\lambda_j(k) + \varepsilon y_i(k), \tag{9}$$

$$P_i(k+1) = \begin{cases} P_i^{\max}, & \lambda^* > \lambda_i^{\max}, \\ \dfrac{\lambda_i(k+1) - \beta_i}{2\alpha_i}, & \lambda_i^{\min} \le \lambda^* \le \lambda_i^{\max}, \\ P_i^{\min}, & \lambda^* < \lambda_i^{\min}, \end{cases} \tag{10}$$

$$y_i(k+1) = \sum_{j=1}^{n} n_{ij}(k) y_j(k) - \big(P_i(k+1) - P_i(k)\big), \tag{11}$$

where y_i is an auxiliary variable which denotes the estimated local power mismatch and ε is a sufficiently small positive constant named learning gain. $\lambda_i^{\max} = ((P_i^{\max} - \beta_i)/2\alpha_i)$ and $\lambda_i^{\min} = ((P_i^{\min} - \beta_i)/2\alpha_i)$. The distributed algorithms (9)–(11) can solve the economic dispatch problem (2) under constraints (3)-(4). Moreover, the following formula can be derived from (11):

$$\sum_{i=1}^{n} y_i(k+1) = \sum_{i=1}^{n}\sum_{j=1}^{n} n_{ij}(k) y_i(k) - \left(P_i(k+1) - P_i(k)\right) \Longrightarrow \sum_{i=1}^{n}\left(y_i(k+1) + P_i(k+1)\right) = \sum_{i=1}^{n}\left(y_i(k) + P_i(k)\right)$$
$$\Longrightarrow \sum_{i=1}^{n}\left(y_i(k+1) + P_i(k+1)\right) = \sum_{i=1}^{n}\left(y_i(0) + P_i(0)\right). \tag{12}$$

It is easy to deduce that $\sum_{i=1}^{n}(P_i(k+1) + y_i(k+1)) = P_D$ will be guaranteed if $\sum_{i=1}^{n}(P_i(0) + y_i(0)) = P_D$. In other words, the equilibrium between power supply and demand can be guaranteed when the algorithm converges only if the selection of initial value satisfies certain conditions.

Most of the existing literature analyzes the impact of attacks based on the assumption of known attack probabilities [17]. In fact, the probability of an attack is generally not fully grasped by the system [21]. In this paper, we explore the effect of DoS attacks without the knowledge of probability distribution information. To this end, a communication network attacked by the DoS attacker is classified into two types:

(1) The communication network topologies keep strongly connected

(2) There exist communication network topologies disconnected, i.e., some units are isolated from the communication network

3.2. Strongly Connected Scenario under DoS Attacks. This section investigates the performance analysis of the dispatch system in the scenario which communication topology graphs of the microgrid keep strongly connected under DoS attacks.

Under this attacks, the communication topology graphs of the microgrid are time-varying, as shown in Figure 1(a), where the solid line represents the normal connected communication links and the dotted line represents compromised communication links caused by DoS attacks. Although the communication network topology no longer maintains a fixed form under DoS attacks, it still keeps strongly connected. In this case, the distributed economic dispatch algorithm (9)–(11) can be carried out to solve the EDP of the microgrid.

Theorem 1. *In the scenario of attacks that the communication network topologies keep strongly connected, if the positive constant ε in algorithm (9)–(11) is sufficiently small and the variables satisfy the initial condition $\sum_{i=1}^{n}(P_i(0) + y_i(0)) = P_D$, then the distributed algorithm is stable, and all the variables converge to the solution of the EDP, i.e.,*

$$\begin{aligned} &\lambda_i(k) \longrightarrow \lambda^*, \\ &P_i(k) \longrightarrow P^*, \\ &y_i(k) \longrightarrow 0, \text{ as } \quad k \longrightarrow \infty, \forall i \in V. \end{aligned} \tag{13}$$

Proof. Use the eigenvalue perturbation approach [18] to analyze the convergence performance of algorithms (9)–(11). Substituting (9) and (10) into (11), we can obtain

$$\lambda(k+1) = M(k)\lambda(k) + \varepsilon y(k), \tag{14}$$

$$y(k+1) = (N(k) - \varepsilon B) y(k) + B(I - M(k))\lambda(k), \tag{15}$$

where λ and y are the column stack vectors of λ_i and y_i, respectively, I is an identity matrix, and $B = \text{diag}\left(1/2\alpha_1 \;\; 1/2\alpha_2 \;\; \ldots \;\; 1/2\alpha_N\right)$. Algorithms (14)–(15) can be written in the following form:

$$\begin{bmatrix} \lambda(k+1) \\ y(k+1) \end{bmatrix} = \begin{bmatrix} M(k) & \varepsilon I \\ B(I - M(k)) & N(k) - \varepsilon B \end{bmatrix} \begin{bmatrix} \lambda(k) \\ y(k) \end{bmatrix}. \tag{16}$$

Define $C(k) = \begin{bmatrix} M(k) & \mathbf{0} \\ B(I - M(k)) & N(k) \end{bmatrix}$, $\Delta = \begin{bmatrix} \mathbf{0} & I \\ \mathbf{0} & -B \end{bmatrix}$, and updating matrix $D(k) = \begin{bmatrix} M(k) & \varepsilon I \\ B(I - M(k)) & N(k) - \varepsilon B \end{bmatrix}$ $= C(k) + \varepsilon\Delta$.

The updating matrix $D(k)$ in (16) can be regarded as $C(k)$ perturbed by $\varepsilon\Delta$. It is easy to deduce that $C(k)$ is a lower triangular block matrix with the matrices $M(k)$ and $N(k)$ in its diagonal line. As a result, the eigenvalues of $C(k)$ include that of matrices $M(k)$ and $N(k)$. According to Lemma 1, it can be obtained that $C(k)$ has two eigenvalues $\nu_1(k) = \nu_2(k) = 1$ at any iteration k.

Assume $z_i(k)$ and $x_i(k)$ are left and right eigenvectors of matrix $C(k)$ with respect to eigenvalues ν_i, and they satisfy the following conditions:

$$\begin{aligned} z_i^T C(k) &= z_i^T \nu_i(k), \\ C(k) x_i(k) &= \nu_i(k) x_i(k), \\ z_i^T x_i(k) &= 1, \\ z_i^T x_j(k) &= 0. \end{aligned} \tag{17}$$

According to formula (17), it is not hard to find that $z_1^T(k) = \left[\mathbf{1}^T B \;\; \mathbf{1}^T\right]$ and $z_2^T(k) = \left[u^T(k) \;\; \mathbf{0}\right]$ both are the left eigenvectors of $C(k)$; $x_1(k) = \left[\mathbf{0}^T \quad \xi(k)^{TT}\right]$ and $x_2(k) = \left[\mathbf{1}^T \;\; -\eta\xi(k)^T\right]^T$ both are the right eigenvectors of $C(k)$ associated with eigenvalues $\nu_1(k) = \nu_2(k) = 1$, where $\eta = \sum_{i=1}^{N}(1/2\alpha_i)$ is defined. If $\varepsilon \neq 0$, the eigenvalues of $C(k)$ is perturbed by $\varepsilon\Delta$. According to the eigenvalue derivation theory, the partial derivative of ν_i corresponding to ε can be expressed as follows:

$$\frac{\partial \nu_i(k)}{\partial \varepsilon} = z_i^T(k) \frac{\partial D(k)}{\partial \varepsilon} x_i(k). \tag{18}$$

Thus, it can be deduced that

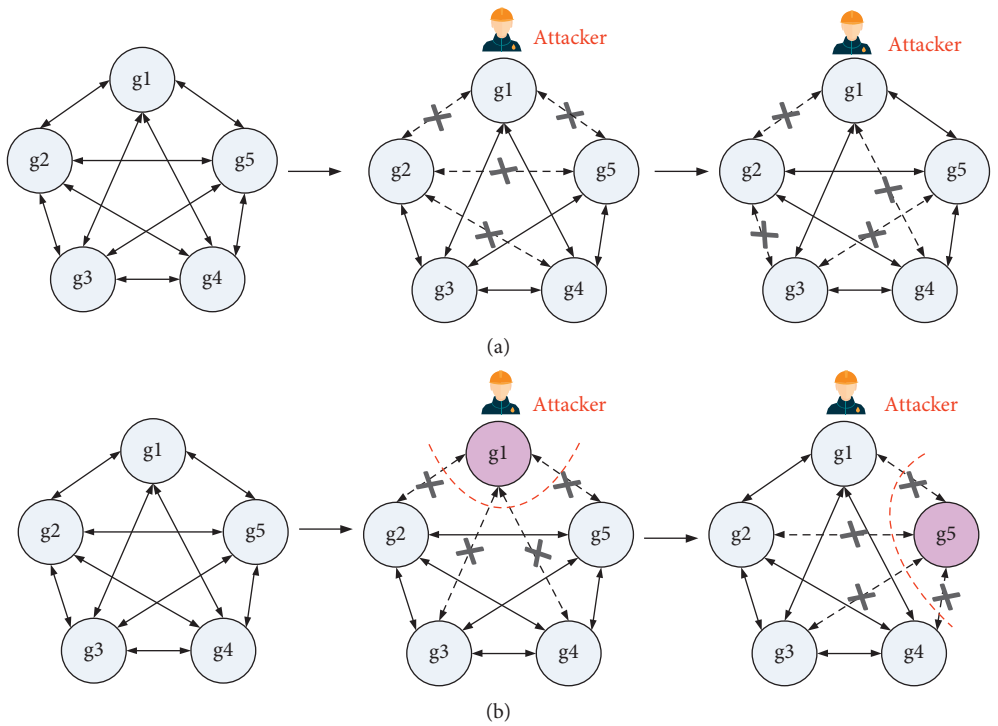


FIGURE 1: Two scenarios caused by DoS attacks. (a) Communication network topologies are still strongly connected. (b) Communication network topologies are disconnected.

$$\frac{\partial \nu_1(k)}{\partial \varepsilon} = z_1^T(k)\frac{\partial D(k)}{\partial \varepsilon}x_1(k) = \begin{bmatrix} \mathbf{1}^T B & \mathbf{1}^T \end{bmatrix}\begin{bmatrix} \mathbf{0} & I \\ \mathbf{0} & -B(k) \end{bmatrix}\begin{bmatrix} \mathbf{0} \\ \xi(k) \end{bmatrix} = 0,$$

$$\frac{\partial \nu_2(k)}{\partial \varepsilon} = z_2^T(k)\frac{\partial D(k)}{\partial \varepsilon}x_2(k) = \begin{bmatrix} \mu^T(k) & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{0} & I \\ \mathbf{0} & -B(k) \end{bmatrix}\begin{bmatrix} \mathbf{1} \\ -\eta\xi(k) \end{bmatrix} \tag{19}$$

$$= -\eta\mu^T(k)\xi(k) < 0.$$

That means $\nu_1(k) = 1$ does not change with ε, and as ε goes up, $\nu_2(k)$ goes down from 1. Therefore, there is a small constant σ_1 satisfying $0 < \varepsilon < \sigma_1$, and $\nu_2(k)$ is restricted to less than 1. Based on the Bauer–Fike theorem [22], there exists an upper bound σ_2 such that $0 < \varepsilon < \sigma_2$, then $\nu_i(k) < 1$, $i = 3, 4 \ldots 2n$, i.e., other eigenvalues of $C(k)$ lie in the open unit disk. Because the eigenvalues of $D(k) = C(k) + \varepsilon\Delta$ depend on ε, given a constant $\sigma = \min\{\sigma_1, \sigma_2\}$, if $0 < \varepsilon < \sigma$, all the eigenvalues of $D(k)$ are less than 1, except for one eigenvalue equal to 1.

It is not hard to deduce that $\begin{bmatrix} \mathbf{1}^T & \mathbf{0}^T \end{bmatrix}^T \in \mathbb{N}^{2n}$ is the right eigenvector of $D(k)$ corresponding to $\nu_1(k) = 1$ at any time k, i.e., $D(k)\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}$. The rest of the eigenvalues are within the open unit disk, and matrix $D(k)$ has $2n$ independent eigenvectors.

It is reasonable to assume that the initial value can be expressed as a linear combination of all the eigenvectors as shown in the following:

$$\begin{bmatrix} \lambda(0) \\ y(0) \end{bmatrix} = c_1\theta_1 + c_2\theta_2 + \cdots + c_{2n}\theta_{2n}, \tag{20}$$

where $\{c_1, c_2, \ldots c_{2n}\}$ is a set of constant and $\{\theta_1, \theta_2, \ldots \theta_{2n}\}$ is a set of eigenvectors of $D(k)$. According to formula (16), we can obtain

$$\begin{bmatrix} \lambda(1) \\ y(1) \end{bmatrix} = D(0)\begin{bmatrix} \lambda(0) \\ y(0) \end{bmatrix} = \nu_1 c_1\theta_1 + \nu_2 c_2\theta_2 + \cdots + \nu_{2n}c_{2n}\theta_{2n}. \tag{21}$$

Thus, it can be deduced that

$$\begin{bmatrix} \lambda(k) \\ y(k) \end{bmatrix} = \prod_{p=1}^{k} D(k-p)\begin{bmatrix} \lambda(0) \\ y(0) \end{bmatrix}. \tag{22}$$

According to the above analysis, because all the eigenvalues of $D(k)$ are less than 1, except for one eigenvalue equal to 1, and the vector $\begin{bmatrix} \mathbf{1}^T & \mathbf{0}^T \end{bmatrix}^T$ corresponds to the eigenvalue 1. It can be conducted that

$$\lim_{k\longrightarrow+\infty}\begin{bmatrix}\lambda(k)\\ y(k)\end{bmatrix}=\operatorname{span}\begin{bmatrix}1\\ 0\end{bmatrix}. \tag{23}$$

As $k\longrightarrow\infty$ and $y_i(k)\longrightarrow 0$, $\sum_{i=1}^{n}(P_i(k)+y_i(k))=\sum_{i=1}^{n}(P_i(0)+y_i(0))$ is always satisfied. Because the initial condition $\sum_{i=1}^{n}(P_i(0)+y_i(0))=P_D$ is guaranteed, it can be derived that the power supply-demand balance of the system is satisfied ultimately. Thus, Theorem 1 can be obtained.

It worth noting that the research in this paper is aimed at day-ahead economic dispatch of the smart grid, and our concern is whether the algorithm can converge to the optimal solution of the EDP rather than the convergence time. This is also the focus of another attack scenario we discuss in Section 3.3. According to Theorem 1, it is not very difficult to obtain that the distributed algorithm executed in the communication environment with DoS attacks still converges to the point which corresponds to the optimal operation point of the microgrid. Through the above analysis, we can get the following inference that, in this scenario of attack that keeps the communication topologies still strongly connected, the DoS attack shows little benefit to the degradation of economic performance for the microgrid. It is a situation that is unwilling to accept by the DoS attacker, whose intent is to disrupt the economic operation of the microgrid as much as possible. □

3.3. Disconnected Scenario under DoS Attacks. This section investigates the performance analysis of the dispatch system in the scenario that communication topology graphs of the microgrid are disconnected under DoS attacks.

If the DoS attacker concentrates on disconnecting all the communication links of a generator, it will briefly prevent the generator from communicating with its neighbors. Under these attacks, at least one generator is isolated from the communication network. It means the time-varying matrix $A(k)$ has the rows and columns in which all the elements are zero, i.e., the in-degree and out-degree of some vertexes in the communication topology graph are zero. For the convenience of analysis, it is necessary to reduce the dimension of $A(k)$ to obtain $A'(k)$ by removing the rows and columns whose elements are all equal to 0. The corresponding matrices $M(k)$ and $N(k)$ are transformed to $M'(k)$ and $N'(k)$, respectively.

In this scenario, all the generators can be divided into two types, misbehaving generators and well-behaving generators, respectively. Define $A^n(k)\subseteq V$ as the set of misbehaving generators at iteration k, i.e., the units isolated from communication network, and $N_n(k)=|A^n(k)|$ is the number of misbehaving generators. It is worth noting that $A^n(k)$ presents isolated generators only due to DoS attacks in unreliable communication network, and isolation of units caused by other abnormal factors is beyond the scope of this paper. If $i\notin A^n(k)$, the generator i is involved in economic dispatch as a well-behaving generator, which updates its output power and incremental cost normally; otherwise, the generator i is a misbehaving generator, which maintains its output power and incremental cost until it is reconnected to the communication network after the disappearance of DoS attacks on it.

Under these attacks, algorithms (9)–(11) can be rewritten as the following dynamics (24)–(26):

$$\lambda_i(k+1)=\begin{cases}\displaystyle\sum_{j\notin A^n(k)} m_{ij}'(k)\lambda_j(k)+\varepsilon y_i(k), & i\notin A^n(k),\\ \lambda_i(k), & i\in A^n(k),\end{cases} \tag{24}$$

$$P_i(k+1)=\begin{cases}P_i^{\max}, & \lambda_i>\lambda_i^{\max},\\ \dfrac{\lambda_i(k+1)-\beta_i}{2\alpha_i}, & \lambda_i^{\min}\le\lambda_i\le\lambda_i^{\max},\\ P_i^{\min}, & \lambda_i<\lambda_i^{\min},\end{cases} \tag{25}$$

$$y_i(k+1)=\begin{cases}\displaystyle\sum_{j\notin A^n(k)} n_{ij}'(k)y_j(k)-\left(P_i(k+1)-P_i(k)\right), & i\notin A^n(k),\\ 0, & i\in A^n(k),\end{cases} \tag{26}$$

$$\sum_{j\notin A^n(k)}P_i^{\min}\le P_D-\sum_{j\in A^n(k)}P_i(k)\le\sum_{j\notin A^n(k)}P_i^{\max}. \tag{27}$$

Theorem 2. *In the scenario of attacks where there exist the communication network topologies disconnected, if the variables satisfy the initial condition* $\sum_{i=1}^{n}(P_i(0)+y_i(0))=P_D$, *the learning gain* ε *in algorithms (24)–(26) is sufficiently small, and there exist enough output power satisfying total power demand, as shown in formula (27); then, the algorithm is stable and the variables of well-behaving generators converge to the suboptimal solution of the EDP, i.e.,*

$$\begin{cases} \lambda_i(k) \longrightarrow \lambda', & i \notin A^n(k), \\ \lambda_i(k) \longrightarrow \lambda_i, & i \in A^n(k), \end{cases}$$

$$\begin{cases} P_i(k) \longrightarrow P', & i \notin A^n(k), \\ P_i(k) \longrightarrow P_i, & i \in A^n(k), \end{cases} \tag{28}$$

$$y_i(k) \longrightarrow 0, \text{ as } \quad k \longrightarrow \infty, \forall i \in V,$$

where $\lambda' \neq \lambda^*$ and $P' \neq P^*$ denote the suboptimal incremental cost vector and output power vector of well-behaving generators, respectively, and λ_i and P_i denote the incremental cost and output power of the misbehaving generator i, respectively. Note that convergence results shown in (28) depend on the attack strategies chosen by the DoS attacker.

Proof. The proof is similar with Theorem 1. The process is omitted here for convenience. □

Remark 1. The attacker deliberately causes algorithms (24)–(26) to be executed on the premise that formula (27) is satisfied in the microgrid. As a result, the algorithm converges to a stable but not optimal point as shown in Theorem 2, the attacks will be considered online stealthily [23]. The attacks are especially hard to detect by the system, and they can keep stealthy and damage the performance of the microgrid.

It worth noting that misbehaving generators isolated from the communication network will continue to supply electricity to local demand because the isolation action caused by DoS attacks cannot directly disconnect the transmission lines of the misbehaving generators in the physical network. Under this scenario of attack, the distributed algorithm cannot be executed consistently due to the absence of the real updating information of the misbehaving units. According to Theorem 2, it is not very difficult to obtain that the distributed algorithm converges to the other convergence point, which is not equal to the optimal operation point of the microgrid. The conclusion can be drawn that, in this attack scenario which make the communication topology graphs disconnected, DoS attacks can interfere with the economic dispatch process of the microgrid to degrade the economic performance of the system. Through the above analysis, this scenario of attacks is a benefit to the DoS attacker. Only if DoS attacks isolate some generators from the communication network can such attacks will be effective for the destruction of economic operation of the microgrid.

Based on the analysis of the effect of attacks on the EDP, the problem that the authors are interested in is to find the largest degradation of economic performance of the microgrid under DoS attacks. In order to quantify the effect of attacks, the change of incremental cost in the attacks' situation compared to the ideal situation is exploited.

4. Attack Strategy Formulation

4.1. Attack Capacity Formulation. According to the analysis in Section 3.1, DoS attacks can occur at any communication moment and communication link. In practice, the launch of DoS attacks cannot be unrestrained, which is reflected in the constrained attack frequency and attack duration [24]. On the one hand, it is due to the inherent characteristics of DoS attacks, and on the other hand, several technologies of the power system can effectively resist jamming attacks [25]. As a consequence, the DoS attacker needs to formulate an optimal attack strategy on the premise of satisfying its own resource constraints to maximize the attack efficiency and attack effect.

We define the total attack resource as Λ, and $\delta_n(k)$ is the attack level at iteration k, which represents the proportion of isolated units to the total number of units, i.e., $\delta_n(k) = (N_n(k)/n)$, $0 \le \delta_n(k) \le 1$. It implies that the more generators that are isolated due to DoS attacks, the higher the attack level the attacks have. Define $\phi = \{\phi_k\}$, $k \in \{0, 1, \ldots\}$ as the attack decision vector in which ϕ_k is the decision of the attacker at iteration k, and $\phi_k = 1$ indicates that the attacker decides to jam the wireless channel; otherwise, $\phi_k = 0$. Define $\rho_k(\phi_k, \delta_k^a(k))$ as the indicator function to represent whether the attacker implements DoS attacks successfully, and the indicator function can be formulated as follows:

$$\rho_k(\phi_k, \delta_n(k)) = \begin{cases} 1, & \phi_k = 1, \delta_n(k) \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{29}$$

4.2. Optimal Attack Strategy. The definition of the incremental cost in (5) implies that incremental cost is the value of the operation cost changing with the generated power. The incremental cost can be regarded as an indicator of economic performance for the power system:

$$\arg\max_{A^n(k)} \sum_{k=k_1}^{k_2} \rho_k \left| \lambda' - \lambda^* \right| = \arg\max_{A^n(k)} \sum_{k=k_1}^{k_2} \rho_k \left| \frac{P_D - \sum_{i \in A^n(k)} P_i(k) + \sum_{i \notin A^n(k)} (\beta_i / 2\alpha_i)}{\sum_{i \notin A^n(k)} (1/2\alpha_i)} - \lambda^* \right|. \tag{30}$$

Thus, the change of incremental cost in the attacks' situation compared to the ideal situation can be used for a metric to quantify the effect of attacks. On this basis, the problem of designing an optimal attack strategy can be formulated as an optimization problem, for which the optimal solution can be obtained mathematically. This optimization problem aims to maximize the incremental cost deviation. At the same time, in order to ensure the DoS

attacks are online stealthy, the attacker should ensure that all the units can meet the power balance requirement after some units are isolated. Furthermore, the DoS attacker needs to formulate an optimal attack strategy on the premise of satisfying its own resource constraints. In the iteration interval $k \in [k_1, k_2]$, the formulation of an optimal DoS attacks' strategy can be transformed to the optimization problem as follows:

$$\text{s.t.} \sum_{k=k_1}^{k_2} \phi_k \delta_n(k) \le \Lambda, \quad \text{and (27)}. \tag{31}$$

The design of optimal strategy is completely reasonable because the purpose of the DoS attacks against economic dispatch process is to violate the economy principle and result in the largest degradation of the economic performance of the microgrid. As shown in previous analysis, λ^* corresponds to the optimal economic dispatch results without DoS attacks and λ' reflects the solution of the EDP under DoS attacks, and the value of λ' is related to the attack strategies chosen by the attacker. Therefore, the more λ' deviates from λ^*, the worse economic performance of the system can be obtained, and the objective function can be formulated as (30). Moreover, if the attack strategy selected violates the power supply-demand principle in (27), the attacks will be detected easily by the monitoring mechanism in the system, which goes against the intention of the attacker to hide its behaviors. Formula (31) considers the constraint of the attack resources from the practical point of view, and the formulation of attack strategy must be limited by it.

In this paper, the static attack problem is investigated, and the optimal attack strategy is decided before the attack actions begin. However, the solution of the above objective function is inaccessible because the output power of isolated generators is unknown at that time. There are only some static parameters available for the attacker to formulate the attack strategies; thus, we put forward the following assumptions about DoS attacks and reformulate the optimization model according to them.

Assumption 1. Attacker can collect pivotal information of the microgrid such as operation cost parameters of generators, power upper and lower limits, and the total power demand. The information can be obtained by monitoring normal process of economic dispatch and collecting the relevant information published by authoritative organizations or manufacturers.

Assumption 2. For the objective to maximize the change of incremental cost in the attacks' situation compared to the ideal situation, the attacker tends to attack the unit when its output power is the lower limit, which increases the burden on well-behaving generators and their incremental cost value.

Based on the assumption of attack pattern and objective shown above, the attacker pursues an expression of the optimal attack strategy X^*. Reformulate (27) and (30), $\sum_{i \in A^n(k)} P_i$ can be replaced by $\sum_{i \in A^n(k)} P_i^{\min}$, and the optimization problem can be reformulated as follows:

$$X^* = \arg\max_{A^n(k)} \sum_{k=k_1}^{k_2} \rho_k \left| \lambda' - \lambda^* \right| = \arg\max_{A^n(k)} \sum_{k=k_1}^{k_2} \rho_k \left| \frac{P_D - \sum_{i \in A^n(k)} P_i^{\min} + \sum_{i \notin A^n(k)} (\beta_i / 2\alpha_i)}{\sum_{i \notin A^n(k)} (1/2\alpha_i)} - \lambda^* \right|, \tag{32}$$

$$\text{s.t.} \sum_{i \notin A^n(k)} P_i^{\min} \le P_D - \sum_{i \in A^n(k)} P_i^{\min} \le \sum_{i \notin A^n(k)} P_i^{\max}, \quad \text{and (31)}, \tag{33}$$

where the parameters' information of (31)–(33) is available in advance, and it means the optimal attack strategy can be decided before the attack actions begin.

Remark 2. According to the above analysis, in the nonideal communication environment with DoS attacks, the incremental cost λ' obtained by the calculation is related to the selected attack strategy. Because the EDP of the microgrid (2)–(4) is a convex optimization problem, the incremental cost is positively correlated with its corresponding output of the generator, which means that the greater the deviation of λ' and λ^* obtained under different attack strategies, the higher the operation cost of the system is. The above optimal attack strategy models (31)–(33) are essential to increase the operating cost of the system, thereby destroying the economic performance of the microgrid.

Based on the analysis of the optimal attack strategy formulation, an algorithm used to obtain this optimal strategy needs to be developed.

4.3. Algorithm to Obtain the Optimal Attack Strategy. The above optimization problem is not a simple convex optimization problem, and it cannot be directly solved by analytical methods or CVX toolbox in MATLAB [26]. As a result, this paper proposes an enumerating-search algorithm to solve the above optimization problem, which is described in Algorithm 1.

The essence of Algorithm 1 is to use the recursive process to exhaustively search for the optimal solution, i.e., the optimal attack strategy which damages the economic performance of the microgrid to the greatest extent. Firstly, enumerate all the possible attack schemes satisfying constraints (31) and (33), where the attack

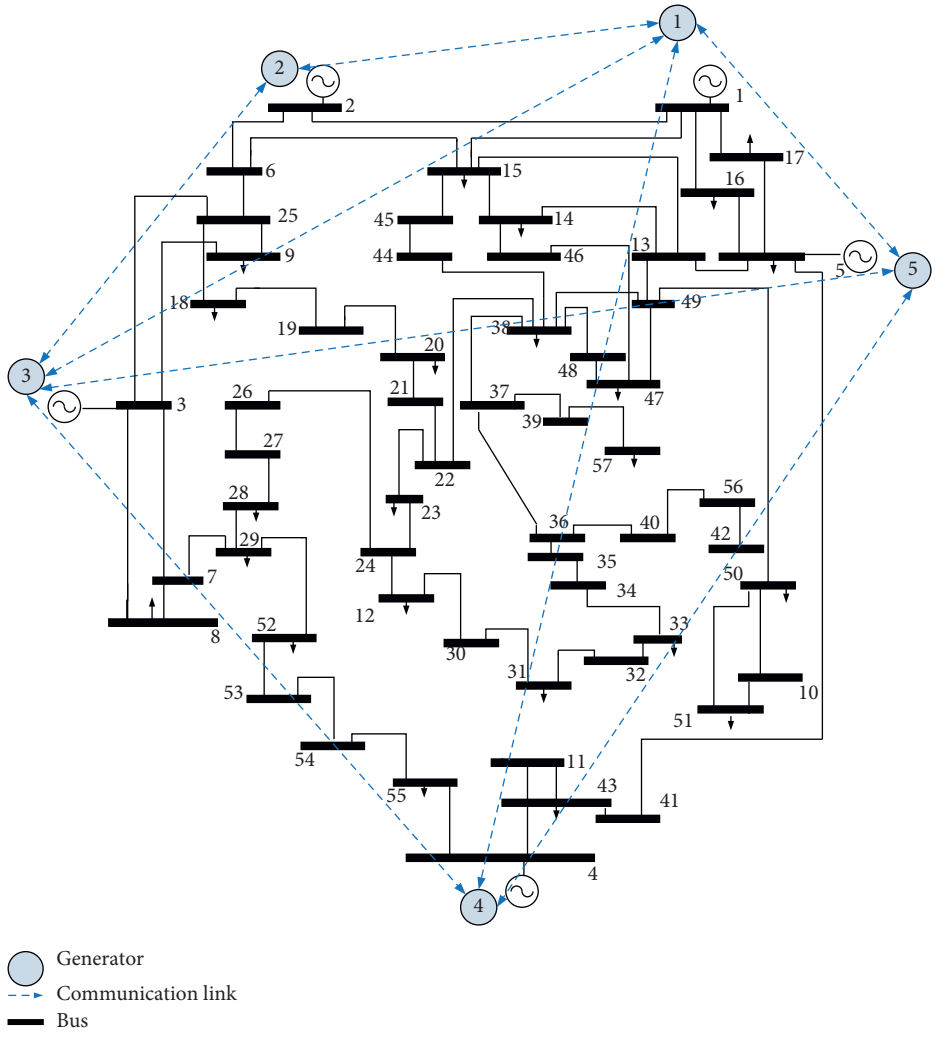


Figure 2: IEEE 57-bus system.

Table 1: Parameters of generators.

α_i	β_i	γ_i	$P_i^{\min}$ (MW)	$P_i^{\max}$ (MW)
0.001562	7.92	561	150	600
0.00194	7.85	310	100	400
0.00482	7.8	78	50	200
0.00234	7.85	300	100	300
0.00312	7.88	100	50	300

Table 2: Initialization.

$\lambda_i(0)$ (\$/MWh)	$P_i(0)$ (MW)	$y_i(0)$ (MW)
8.8572	300	100
8.82	250	−100
8.764	100	50
7.85468	200	−50
7.88624	100	50

schemes are represented by the sequence number of the isolated units caused by each attack action. Then, we compute the deviation value of incremental cost corresponding to each attack scheme. By comparing all the deviation value, the attack scheme corresponding to the maximum deviation value is pursued by the DoS attacker. Finally, combine this optimal attack scheme with the attack decision vector; the attacker can decide the optimal attack strategy. It is worth noting that, for developing the optimal attack strategy before launching attacks so that the DoS attacks achieve the best effect of attacks, the attacker needs to further decide the attack moments, which can be represented by the attack decision vector ϕ [17]. This part is not the focus of the EDP for the microgrid, so it will not be detailed here.

5. Simulation Examples

Numerical simulation is a common method to verify theoretical analysis results as well as the effectiveness of the algorithm proposed. In this section, we firstly chose the IEEE 57-bus system to simulate the microgrid to compare the EDP under the different communication situations. Then, on the IEEE 39-bus system, we discuss the effect of DoS attacks adopting various attack strategies on economic performance of the microgrid.

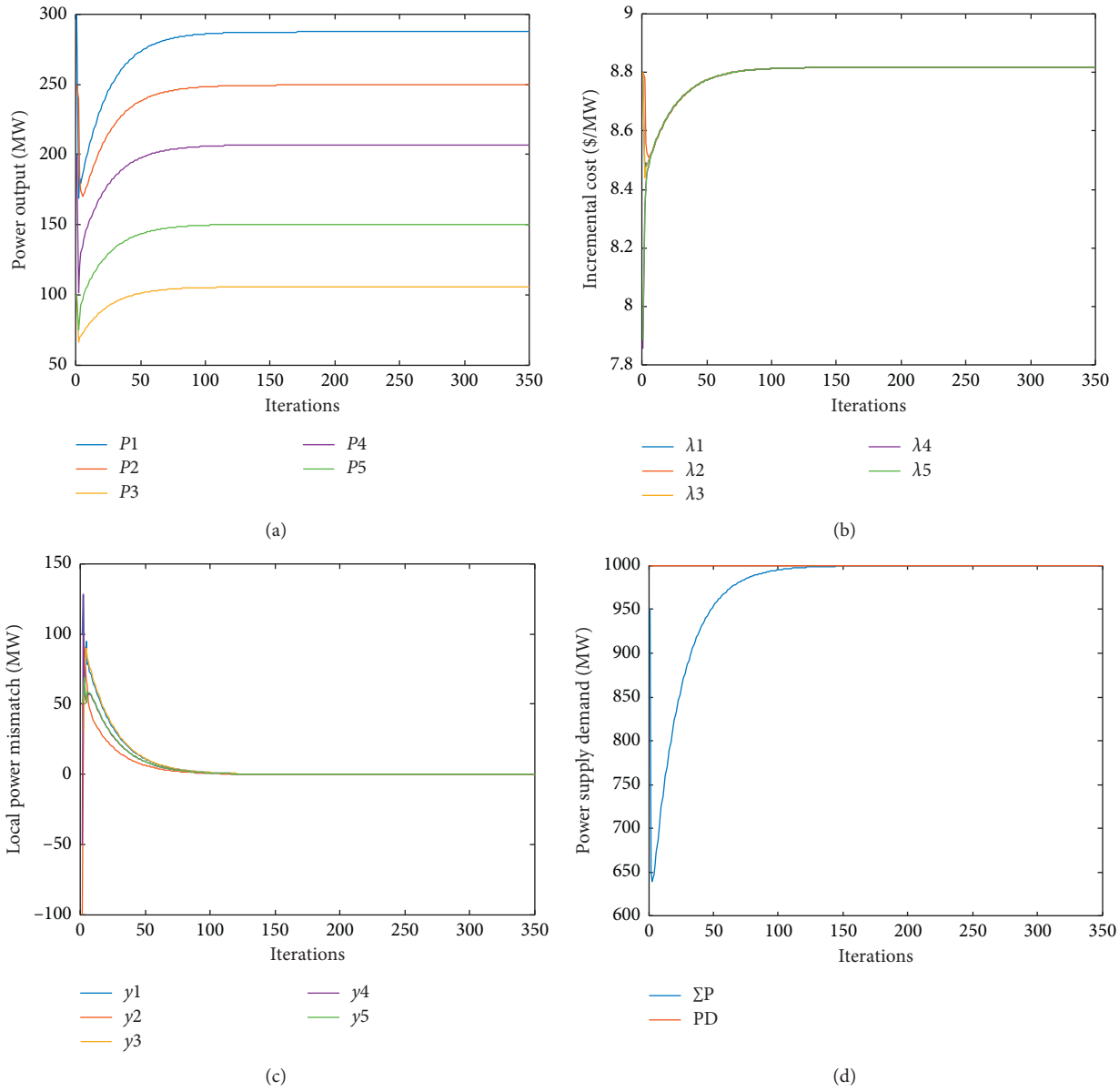


FIGURE 3: The convergence results of the dispatch system under ideal communication conditions. (a) The power output of each generator. (b) The incremental cost of each generator. (c) Auxiliary variable y of each generator. (d) Total power output and load demand.

5.1. Case 1: The EDP under Ideal Communication Conditions. There is an IEEE 57-bus system for simulation, and the communication network corresponding to the physical power network is shown in Figure 2. This test system includes five generators, whose parameters and initial state values are listed in Tables 1 and 2, respectively. The total power demand of the system is set to 1000 MW.

In this case, under the ideal communication conditions without DoS attacks, the EDP can be solved with the algorithm proposed in [20]. In the iteration interval $k \in [0, 350]$, the evolution of the output power P_i, incremental cost λ_i, the total power generated $\sum P$, and power demand P_D is shown in Figure 3. Ultimately, the incremental cost λ_i, $i = 1, \ldots, 5$, converges to a common value $\lambda^* = 8.8183$ \$/MWh; meanwhile, the output power $P_1^* = 287.5440$ MW, $P_2^* = 249.5587$ MW, $P_3^* = 105.6315$ MW, $P_4^* = 206.8991$ MW, and $P_5^* = 150.3666$ MW, all the generators are operating within their generation constraints, and also the total power demand is satisfied by these power output. The convergence results of the distributed algorithm are the same as that calculated by the centralized algorithm in [27]. The total operation cost is 9692.8 \$/h, which is the minimum operation cost of the system.

According to the above analysis, the optimal solution of the EDP can be solved through the distributed economic dispatch algorithm under ideal communication conditions. This case can be used as a comparison of the subsequent simulation cases.

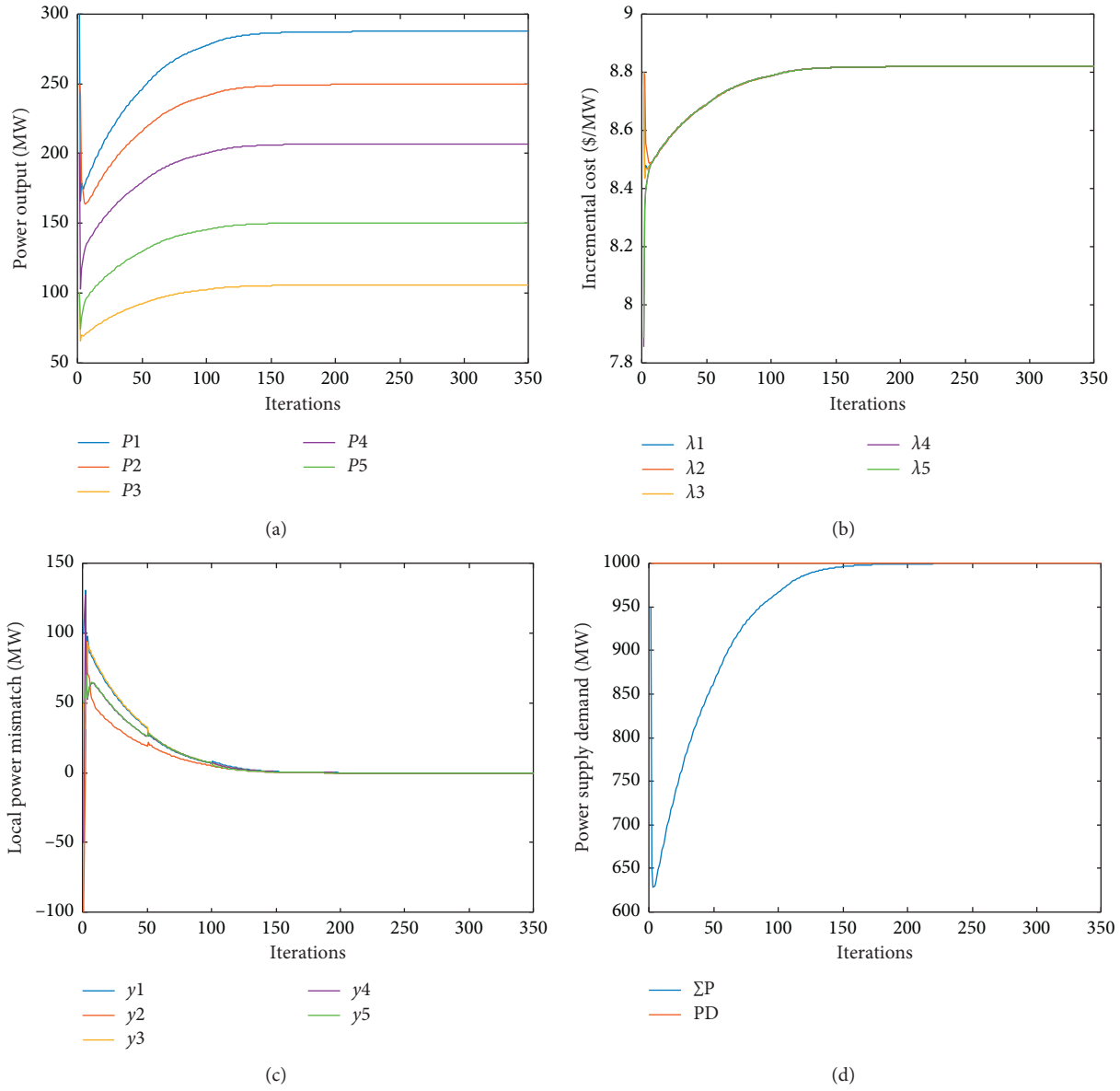


Figure 4: EDP with strongly connected topology under DoS attacks. (a) The power output of each generator. (b) The incremental cost of each generator. (c) Auxiliary variable y of each generator. (d) Total power output and load demand.

5.2. Case 2: The EDP with Strongly Connected Communication Topology under DoS Attacks. Simulation case is performed with the same system model and parameters as those in Section 5.1. In this case, we consider the scenario that communication topology graphs keep strongly connected under DoS attacks. Assuming the communication topology changes two times resulted by DoS attacks, the attacker launches DoS attacks firstly in the communication link (1,3) at iteration $k = 50$ until $k = 100$; then, the attacker launches DoS attacks in the communication link (1,4) at iteration $k = 100$ until $k = 350$.

In this case, the EDP can be solved with algorithms (9)–(11). In the iteration interval $k \in [0, 350]$, the evolution of the output power P_i, incremental cost λ_i, the total power generated $\sum P$, and power demand P_D is shown in Figure 4. Ultimately, the incremental cost $\lambda_i, i = 1, \ldots, 5$, converges to a common value $\lambda^* = 8.8183$ \$/MWh and the output power $P_1^* = 287.5440$ MW, $P_2^* = 249.5587$ MW, $P_3^* = 105.6315$ MW, $P_4^* = 206.8991$ MW, and $P_5^* = 150.3666$ MW. It can be concluded that the final convergence results in this case are the same as the results of Case 1.

According to Figure 4, it can be observed that the distributed algorithms (9)–(11) can finally converge to the optimal solution of EDP, although the existences of DoS attacks interfere with the convergence process of the algorithm. As a result, the correctness of the conclusion in Section 3.2 is verified, i.e., this attack scenario where the communication network still keeps strongly connected shows little benefit to the degradation of economic performance of the microgrid.

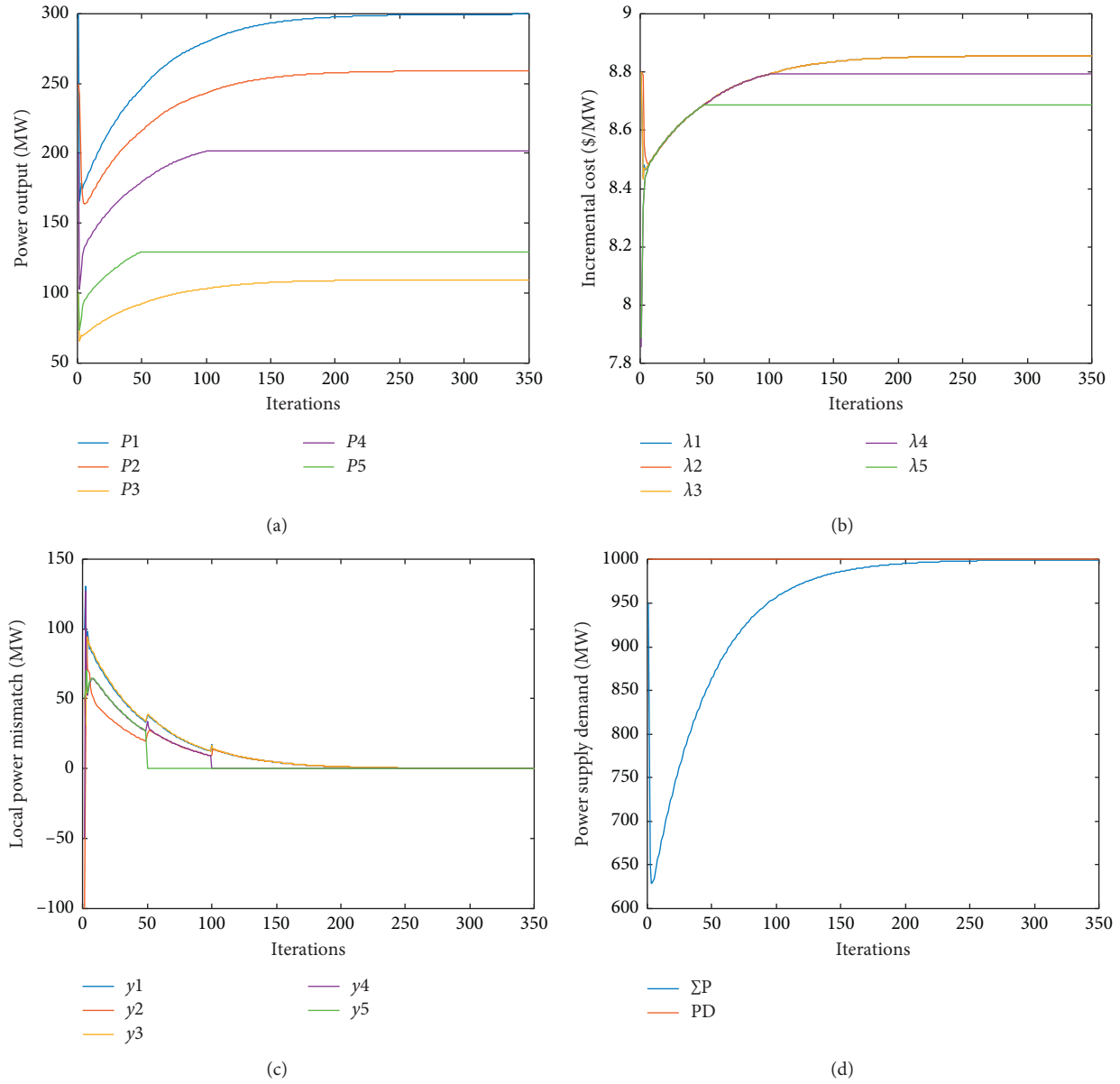


Figure 5: EDP with disconnected topology under DoS attacks. (a) The power output of each generator. (b) The incremental cost of each generator. (c) Auxiliary variable y of each generator. (d) Total power output and load demand.

5.3. Case 3: The EDP with Disconnected Communication Topology under DoS Attacks. Simulation is performed with the same system model and parameters as those in Section 5.1. In this case, we consider the scenario that there exist communication topology graphs disconnected under DoS attacks. If all the communication paths of a generator are attacked for a certain period of time, the generator is briefly isolated from the communication network. It is important to note that the generator will reconnect to the communication network when the attacks disappear.

Assuming the communication topology changes two times because of DoS attacks, the attacker launches DoS attacks on generator 5, firstly, at iteration $k = 50$ until $k = 100$; then, the attacker launches DoS attacks on generator 4 and 5 at iteration $k = 100$ until $k = 350$. In this scenario, these two generators are misbehaving generators because of the DoS attacks on them, which lead to the disconnected communication topology graphs.

In this case, algorithms (24)–(26) are carried out to solve the EDP. The evolution of the output power P_i, incremental cost λ_i, the total power generated $\sum P$, and power demand P_D is shown in Figure 5. Under these DoS attacks, misbehaving generator 5 is isolated from the communication network at the iteration interval $k \in [50, 350]$, and misbehaving generator 4 is isolated at the iteration interval $k \in [100, 350]$, while the output power of the well-behaving generators can still satisfy the power demand remained. The incremental cost $\lambda_i, i = 1, 2, 3$, converge to a common value $\lambda' = 8.8562$ \$/MWh, λ_4 converges to $\lambda_4 = 8.7934$ \$/MWh, and λ_5

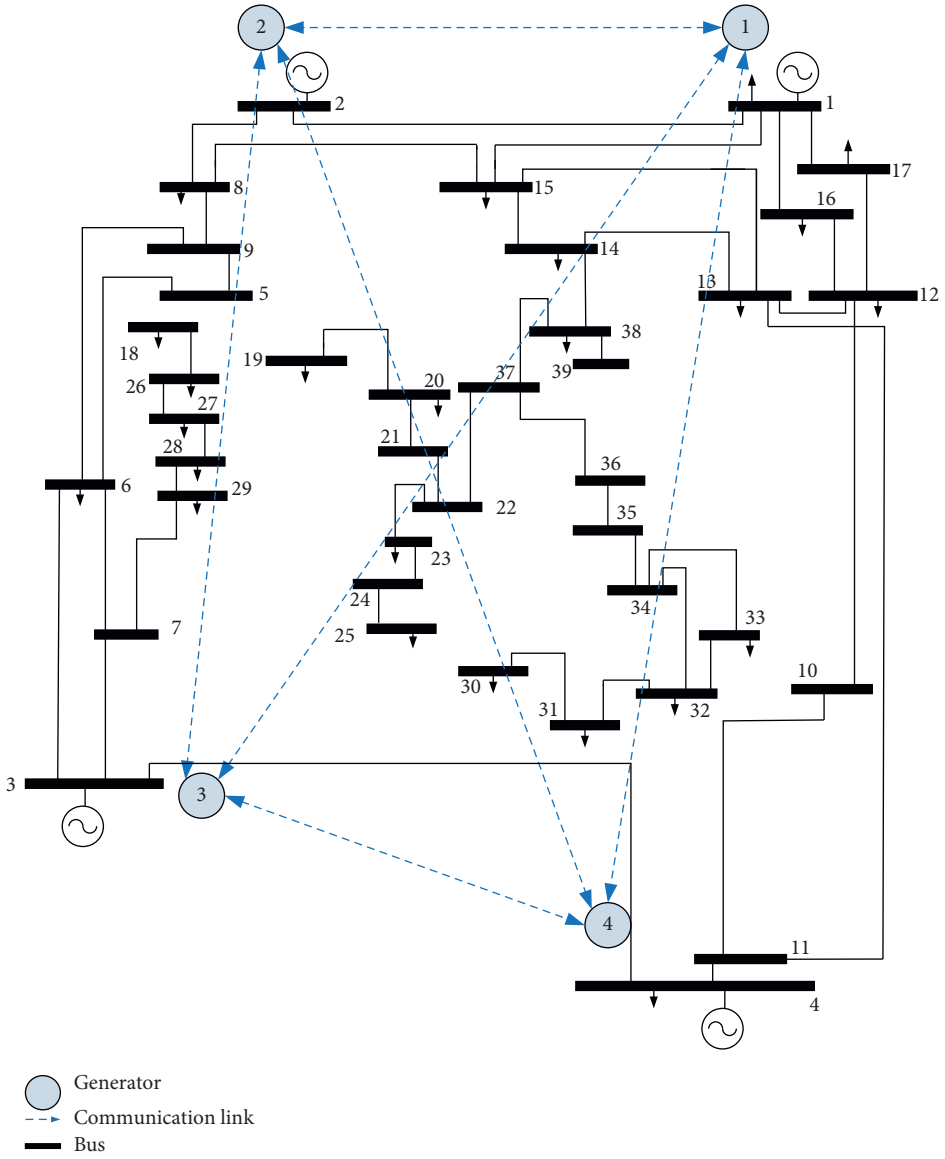


Figure 6: IEEE 39-bus system.

converges to $\lambda_5 = 8.6893$ \$/MW. The ultimate generators' output $P_1' = 299.6912$ MW, $P_2' = 259.3390$ MW, $P_3' = 109.5680$ MW, $P_4' = 201.5769$ MW, and $P_5' = 129.6879$ MW. All the generations are operating within their generation constraints, and the total load requirement is satisfied ultimately by generators participating in economic dispatch and the generators isolated. The total operation cost of the system is 9693.5 \$/h, which is higher than that of Case 1, the economic performance of the system declines.

According to Figure 5 and the total operation cost, it can be observed that the distributed algorithms (24)–(26) cannot finally converge to the optimal solution of the EDP, and the economic dispatch process is disrupted by malicious attacks. It implies the correctness of conclusion in Section 3.3 is verified, i.e., the attack scenario that makes the communication network disconnected leads to the suboptimal solution of the EDP, which degrades the economic performance of the microgrid.

5.4. Case 4: The Optimal Attack Strategy. In this case, we discuss the different attack strategies, which can cause varying degrees of damage to economic performance of the microgrid. We chose the IEEE 39-bus system to simulate the microgrid, and its communication network corresponding to the physical power network is shown in Figure 6. This system includes four generators, whose parameters can be chosen in Table 1. The total power load in the microgrid is set to 500 MW. Assuming the total attack energy $\Lambda = 1$ and attack decision vector $\phi = \{0 \ \ldots \ 0 \ 1 \ 0 \ \ldots \ 0 \ 1 \ 0 \ \ldots \ 0\}$, it is easy to deduce that $\delta_n(k)$ must be smaller than 1/2, which means the number of generators isolated caused by each attack action cannot exceed two for the communication network in Figure 6. Then, the set of attack strategies can be obtained according to constraints (31) and (33).

For the communication network in this case, we can enumerate the set of all attack schemes

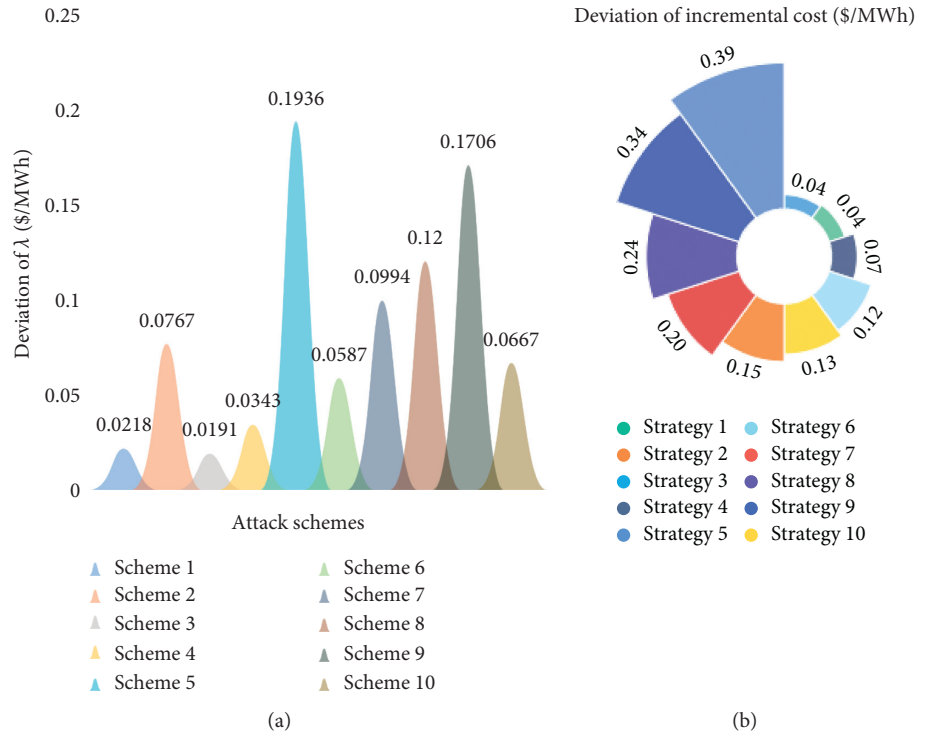


FIGURE 7: The comparison for different attack strategies. (a) Deviation of incremental cost in the attacks' situation compared to the ideal situation. (b) The effect of attacks under different attack strategies.

$X = \{X_1, X_2, \ldots X_{10}\}$, which include $X_1 = \{1\}$, $X_2 = \{2\}$, $X_3 = \{3\}$, $X_4 = \{4\}$, $X_5 = \{1, 2\}$, $X_6 = \{1, 3\}$, $X_7 = \{1, 4\}$, $X_8 = \{2, 3\}$, $X_9 = \{2, 4\}$, and $X_{10} = \{3, 4\}$; an attack scheme is represented by the serial numbers of the misbehaving generators caused by each attack action. Establishing the objective function and constraints as (31)–(33), the optimization problem can be solved by Algorithm 1, and the calculation results are shown in Figure 7(a). According to Figure 7(a), if $X_5 = \{1, 2\}$ is adopted, the deviation of incremental cost will be maximized. It means the execution of the optimal attack strategy X^* is to repeat the attack scheme $X_5 = \{1, 2\}$ at each attack moment. Combining the attack decision vector ϕ and attack schemes, different attack strategies can be obtained. The comparison of different attack strategies is shown in Figure 7(b), from which we can easily sum up that optimal attack strategy obtained by using Algorithm 1 indeed causes the most degradation of economic performance compared with other attack strategies. As a result, it is feasible to carry out the proposed enumerating-search algorithm to obtain the optimal attack strategy.

6. Conclusion

This paper considers an optimal attack strategy against the economic dispatch for the microgrid. Firstly, according to the operating principle of DoS attacks, the unreliable communication network under DoS attacks can be classified into two scenarios, i.e., disconnected and strongly connected scenarios. The theoretical analysis results show that economy performance of the microgrid is damaged in the former case. On this basis, the optimal DoS attack strategy considering the resource constraints of the attacker is formulated to maximize the effect of attacks, which can be obtained by using the enumerating-search algorithm. Finally, the correctness of the theoretical analysis results with respect to the effect of attacks is verified by the simulations on the IEEE 57-bus system, and the effectiveness of the proposed enumerating-search-based optimal attack strategy is illustrated through the simulations on the IEEE 39-bus system. This paper delivers the message that, besides the efforts of designing novel distributed economic dispatch algorithms to guarantee the economic operation for the microgrid, it is equally important to protect the economic dispatch process from malicious attacks and evade potential economic losses.

For the future work, how to build the rational model of DoS attacks is the key to solving the EDP of the energy system under DoS attacks. Furthermore, we will synthetically consider the impact of other attacks on the EDP, such as false data injection attacks and replay attacks. Moreover, research on the dynamic optimal attack strategy disrupting the economic dispatch of the energy system is also the focus of our future work.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported partly by the National Key R&D Program of China under Grant 2018YFA0702200, Fundamental Research Funds for the Central Universities (N2004005), and Science and Technology Project of State Grid Corporation of China (2020YF-19).

References

[1] B. Huang, L. Liu, H. Zhang, Y. Li, and Q. Sun, "Distributed optimal economic dispatch for microgrids considering communication delays," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 8, pp. 1634–1642, 2019.

[2] R. Wang, Q. Sun, W. Hu, Y. Li, D. Ma, and P. Wang, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, p. 7866, 2021.

[3] H. Li, Z. Wang, G. Chen, and Z. Y. Dong, "Distributed robust algorithm for economic dispatch in smart grids over general unbalanced directed networks," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4322–4332, 2020.

[4] W. Hu, C. Ruan, H. Nian, and D. Sun, "Zero-sequence current suppression strategy with common-mode voltage control for open-end winding PMSM drives with common DC bus," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 6, p. 4691, 2021.

[5] J. Zhang, J. Zhang, F. Zhang, M. Chi, and L. Wan, "An improved symbiosis particle swarm optimization for solving economic load dispatch problem," *Journal of Electrical and Computer Engineering*, vol. 2021, Article ID 8869477, 11 pages, 2021.

[6] J. B. Park, Y. W. Jeong, and J. R. Shin, "An improved particle swarm optimization for nonconvex economic dispatch problems," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 156–166, 2020.

[7] A. Cherukuri and J. Cortés, "Initialization-free distributed coordination for economic dispatch under varying loads and generator commitment," *Automatica*, vol. 74, pp. 183–193, 2016.

[8] Z. Fu, X. He, T. Huang, and H. Abu-Rub, "A distributed continuous time consensus algorithm for maximize social welfare in micro grid," *Journal of the Franklin Institute*, vol. 353, no. 15, pp. 3966–3984, 2016.

[9] Y. Li, W. Gao, G. Wei, H. Zhang, and J. Zhou, "A distributed doublenewton descent algorithm for cooperative energy management of multiple energy bodies in energy internet," *IEEE Transactions on Industrial Informatics*, vol. 1, p. 1, 2020.

[10] H. Li and T. Huang, "Convergence analysis of a distributed optimization algorithm with a general unbalanced directed com- munication network," *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 3, pp. 237–248, 2018.

[11] D. Ding, Q.-L. Han, Y. Xiang, X. Ge, and X.-M. Zhang, "A survey on security control and attack detection for industrial cyber-physical systems," *Neurocomputing*, vol. 275, no. 10, pp. 1674–1683, 2018.

[12] E. Hammad, A. Farraj, and D. Kundur, "On cyber-physical coupling and distributed control in smart grids," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4418–4429, 2019.

[13] W. Xu, D. W. C. Ho, J. Zhong, and B. Chen, "Event/self-triggered control for leader-following consensus over unreliable network with DoS attacks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 10, pp. 3137–3149, 2019.

[14] P. Li, Y. Liu, H. Xin, and X. Jiang, "A robust distributed economic dispatch strategy of virtual power plant under cyber-attacks," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4343–4352, 2018.

[15] D. Zhang and G. Feng, "A new switched system approach to leader-Ffollower consensus of eterogeneous linear multiagent systems with DoS attack," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 99, pp. 1–9, 2019.

[16] C. Yang, X. Ren, W. Yang, H. Shi, and L. Shi, "Jamming attack in centralized state estimation," in *Proceedings of the 2015 34th Chinese Control Conference (CCC)*, pp. 6530–6535, Hangzhou, Zhejiang, China, July 2015.

[17] H. Zhang, Y. Qi, J. Wu, L. Fu, and L. He, "DoS attack energy management against remote state estimation," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 383–394, 2018.

[18] Z. Zhang and M.-Y. Chow, "Convergence analysis of the incremental cost consensus algorithm under different communication network topologies in a smart grid," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1761–1768, 2012.

[19] B. Huang, L. Liu, Y. Li, and H. Zhang, "Distributed optimal energy management for microgrids in the presence of time-varying communication delays," *IEEE Access*, vol. 7, pp. 83702–83712, 2019.

[20] S. Yang, S. Tan, and J.-X. Xu, "Consensus based approach for economic dispatch problem in a smart grid," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4416–4426, 2013.

[21] S. Liu, Z. Hu, X. Wang, and L. Wu, "Stochastic stability analysis and control of secondary frequency regulation for islanded microgrids under Random denial of service attacks," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4066–4075, 2019.

[22] F. L. Bauer and C. T. Fike, "Norms and exclusion theorems," *Numerische Mathematik*, vol. 2, no. 1, pp. 137–141, 1960.

[23] C. Zhao, J. He, P. Cheng, and J. Chen, "Analysis of consensus-based distributed economic dispatch under stealthy attacks," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5107–5117, 2017.

[24] C. De Persis and P. Tesi, "Input-to-state stabilizing control under denial-of-service," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 2930–2944, 2015.

[25] W. Xu, K. Ma, W. Trappe, and Y. Zhang, "Jamming sensor networks: attack and defense strategies," *Network IEEE*, vol. 20, no. 3, pp. 41–47, 2006.

[26] D. A. Guimaraes, G. H. F. Floriano, and L. S. Chaves, "A tutorial on the CVX system for modeling and solving convex optimization problems," *IEEE Latin America Transactions*, vol. 13, no. 5, pp. 1228–1257, 2015.

[27] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*, Wiley-Interscience, New York, NY, USA, 1996.

Hindawi
Complexity
Volume 2021, Article ID 6646799, 10 pages
https://doi.org/10.1155/2021/6646799

Research Article

A Brief Overview of Optimal Robust Control Strategies for a Benchmark Power System with Different Cyberphysical Attacks

Bo Hu,[1] **Hao Wang**,[2,3] **Yan Zhao**,[2,3] **Hang Zhou**,[2,3] **Mingkun Jiang**,[2,3] **and Mofan Wei**[2,3]

[1]*State Grid Liaoning Electric Power Supply Co. Ltd.,, Shenyang 110004, China*
[2]*School of Renewable Energy, Shenyang Institute of Engineering, Shenyang 110136, China*
[3]*Key Laboratory of Regional Multi-energy System Integration and Control of Liaoning Province, Shenyang 110136, China*

Correspondence should be addressed to Bo Hu; bohuln@163.com

Received 29 November 2020; Revised 20 January 2021; Accepted 6 February 2021; Published 23 March 2021

Academic Editor: Rui Wang

Security issue against different attacks is the core topic of cyberphysical systems (CPSs). In this paper, optimal control theory, reinforcement learning (RL), and neural networks (NNs) are integrated to provide a brief overview of optimal robust control strategies for a benchmark power system. First, the benchmark power system models with actuator and sensor attacks are considered. Second, we investigate the optimal control issue for the nominal system and review the state-of-the-art RL methods along with the NN implementation. Third, we propose several robust control strategies for different types of cyberphysical attacks via the optimal control design, and stability proofs are derived through Lyapunov theory. Furthermore, the stability analysis with the NN approximation error, which is rarely discussed in the previous works, is studied in this paper. Finally, two different simulation examples demonstrate the effectiveness of our proposed methods.

1. Introduction

With the development of cloud computing, artificial intelligence, and 5th-generation, the power systems regarded as the primary infrastructures in society become typical CPSs [1, 2]. Since there are numerous physical sensors, complex interaction mechanisms, and massive signals in cyber-physical power systems [3], the security of CPSs is inevitably threatened. For example, a large-scale blackout caused by cyberattacks in Ukraine had disrupted the normal lives of many people [4, 5]. Despite there are many advanced control strategies in CPSs, the imperfection of security has not been sufficiently addressed. Meanwhile, the power system composed of distributed energy and multiple loads is multidimensional [6]. It is urgent to further strengthen the security of CPSs.

Generally, the security of CPSs is threatened by attacks from the perception layer, cyberlayer, and decision layer. In particular, the attacks at the perception layer and cyberlayer, known as cyberattacks, severely disrupt the system. In recent years, reliable control strategies against various cyberattacks, such as false data injection attacks, time-delay switch attacks, and denial-of-service attacks, have been presented by many scholars. Denial-of-service attacks, which can jam information transmission channel, are an aggressive threat to CPS security [7–9]. A novel control strategy based on the game theoretic approach was proposed to resist the attacks in discrete systems [7]. Similar to the theory of literature [7], Seo et al. [9] primarily solved jamming attack in the communication between sensor and network, where an adaptive scheduling with energy constraints was presented. Using an evaluation function that quantitatively analyzes the impact of attacks, the optimal attack strategy was investigated under energy constraint in a wireless network, which can maximally destroy the stability of the system [10]. Besides, false data injection attacks, which generally cause state estimation errors, have received widespread attention because these attacks can send inaccurate control signals to the executor [11–14]. Moreover, the critical detection technology for unknown attacks and unpredictable attack areas was proposed in the previous works [9, 15–20]. For instance, considering the characteristics of the network topology and

transmission media, a risk prediction method based on a predictive model was proposed to accurately obtain the characteristics of the physical system, which can judge the fault area in the CPSs [16]. In [20], focusing on undetectable attacks, a dynamic attack detector was proposed.

With the integration of multiple energy sources, the control platform and information transmission are extremely complicated [21]. Thus, other irresistible attacks, sensor, and actuator attacks are a topic of research. For example, in [22], a reliable control with the attack compensator was investigated which can withstand sensor and actuator attacks. In [1], the resilient control strategies were proposed to ensure that the variables converge to the equilibrium point in presence of sensor and actuator attacks.

This paper concentrates on the study of an optimal robust control strategy, where the designed unified control method makes the power system immune to the actuator and sensor attacks. We use optimal control theory, reinforcement learning (RL), and neural networks (NNs) to design the controller under the assumed attacks of multiple characteristics. The main works and contributions can be summarized as follows:

(1) Optimal control theory, RL, and NNs are integrated to address the security issue of a benchmark power system.

(2) A unified way is proposed to deal with the sensor and actuator attacks via the optimal control design.

(3) The stability analysis with the NN approximation error, which is rarely discussed in the previous works, is studied in this paper.

The rest of this paper is arranged as follows.

First, the benchmark power system models with actuator and sensor attacks are formulated. Second, the optimal control issue for the nominal system is investigated, and the state-of-the-art RL methods along with the NN implementations are reviewed. Third, several robust control strategies are proposed for different types of cyberphysical attacks via the optimal control design, and stability proofs are derived through Lyapunov theory. Then, two different simulation examples demonstrate the effectiveness of our proposed methods. Finally, a brief conclusion is given.

2. Problem Statement for Power System

Let us consider the following benchmark power system:

$$\begin{aligned}\Delta\dot{\psi}_f &= -\frac{1}{T_p}\Delta\psi_f + \frac{k_p}{T_p}\Delta\psi_t,\\ \Delta\dot{\psi}_t &= -\frac{1}{T_t}\Delta\psi_t + \frac{1}{T_t}\Delta\psi_g,\\ \Delta\dot{\psi}_g &= -\frac{1}{k_s T_g}\Delta\psi_f - \frac{1}{T_g}\Delta\psi_g + \frac{1}{T_g}u,\end{aligned} \tag{1}$$

where $\Delta\psi_f$, $\Delta\psi_t$, and $\Delta\psi_g$ represent the deviations of frequency, turbine power, and governor position value, respectively; T_t, T_g, and T_p denote the time constants of turbine, governor, and power system, respectively; k_p and k_s represent the gain of power system and the speed regulation coefficient, respectively; and u is the control input.

Let $x = [\Delta\psi_f, \Delta\psi_t, \Delta\psi_g]^T$. The nominal system (1) can be rewritten as

$$\dot{x} = Ax + Bu, \tag{2}$$

where $A = \begin{bmatrix} -(1/T_p) & (k_p & /T_p) & 0 \\ 0 & -(1/T_t) & (1/T_t) & \\ -(1/k_s T_g) & 0 & -(1/T_g) & \end{bmatrix}$ and $B = \begin{bmatrix} 0 \\ 0 \\ (1/T_g) \end{bmatrix}$.

However, the attacks on the system are generally inevitable, which may affect the control performance. System dynamics (2) suffers from the actuator and sensor attacks, which can be, respectively, described by

$$\dot{x} = Ax + B(\overline{u} + t\Lambda), \tag{3}$$

$$\dot{x} = Ax + B\overline{u} + \Lambda, \tag{4}$$

where $\overline{u}$ is the robust control policy, which will be designed later. Λ denotes the system uncertainties. In this paper, we will consider different types of attacks.

Due to the existence of unknown attacks, it is difficult or even impossible to investigate the systems (3) and (4) directly. Inspired by the idea of classical works [23–26], we convert this robust control issue of the systems (3) and (4) into the optimal control problem of the nominal system (2). The main idea is that, with the system data and models, we can first attain the optimal control policy through ADP algorithms. Subsequently, based on the optimal control form, we can develop different robust control strategies for the systems with various attacks.

3. Optimal Control for the Nominal System

Define the performance index function as

$$J(x(0), u) = \int_0^\infty r(x(\tau), u(\tau))\mathrm{d}\tau, \tag{5}$$

where $r(x, u) = x^T Q x + u^T R u$ with positive definite symmetric matrices Q and R. Given the admissible control policy $u(x)$, the value function is expressed as

$$V(x(t)) = \int_t^\infty r(x(\tau), u(x(\tau)))\mathrm{d}\tau. \tag{6}$$

The optimal value function can be defined as

$$V^*(x(t)) = \min_u \left(\int_t^\infty r(x(\tau), u(x(\tau)))\mathrm{d}\tau \right). \tag{7}$$

According to the stationarity condition [27], the optimal control policy is derived by

$$u^*(x) = -\frac{1}{2} R^{-1} B^T \nabla V^*(x), \tag{8}$$

where $\nabla V^*(x) = \partial V^*(x)/\partial x$ and $V^*(x)$ should satisfy the following Hamilton–Jacobi–Bellman (HJB) equation.

$$0 = r\left(x, u^*(x)\right) + \nabla V^{*T}(x)\left(Ax + Bu^*(x)\right). \tag{9}$$

Thus, the key point to obtain the optimal control policy is to solve the HJB equation.

ADP is a powerful tool to solve the optimal control problems. Traditional ADP methods include two iterative algorithms: policy iteration (PI) and value iteration (VI). Afterwards, two noniterative RL methods are developed.

3.1. Online RL method. The aforementioned iterative ADP methods belong to the offline learning field because the value function and control policies are updated with the iteration index. Quite different from offline algorithms, online RL methods [27, 28] do not involve any iteration processes, and the value function and control policies are updated in real time.

3.2. Event Trigger-Based RL Method. In the online RL methods, the update and delivery of information must be continuous, which causes a waste of communication resources. For this phenomenon, the event trigger-based RL methods [29, 30] are developed. Here, the value function and control policies are updated only once when the system state error reaches the set point, which reduces the communication burden.

By using the aforementioned ADP methods, we can obtain the optimal control form of the nominal system, which will be employed in the following sections.

To implement the proposed algorithms, a critic NN and an actor NN are employed to approximate the iterative value function and control policy:

$$\widehat{V}^{(i)}(x) = \phi_c^T(x) W_c^{(i)}, \tag{10}$$

$$\widehat{u}^{(i)}(x) = \phi_a^T(x) W_a^{(i)}, \tag{11}$$

where $\phi_c(x)$ and $\phi_a(x)$ denote the NN activation functions and $W_c^{(i)}$ and $W_a^{(i)}$ represent the NN weights.

Hence, the optimal value function and control policy have NN representation as

$$V^*(x) = \phi_c^T(x) W_c, \tag{12}$$

$$u^*(x) = \phi_a(x) W_a, \tag{13}$$

where W_c and W_a denote the ideal NN weights.

In previous works, the NN approximation error was rarely discussed. In this paper, we attempt to consider its effect in the stability analysis.

In Figure 1, the sensor attacks, tampering the state values collected by sensors, occur in the sensor and communication network. Meanwhile, the actuator attacks, which generally modify the control instructions in actuator, occur between the decision and physical layer. The changed system state and control command can be eliminated by the robust control strategy which is calculated by RL based on the performance index function. Ultimately, the power system can work at the scheduled operating point under the sensor and actuator attacks.

4. Robust Control Strategies for Actuator Attacks

First, let us consider the system (3) with $\Lambda = \Gamma(x)$, where $\|\Gamma(x)\| \le k_d \|x\|$. The robust controller is designed by

$$\overline{u}(x) = u^*(x), \tag{14}$$

where the parameters for generating $u^*(x)$ will be determined later.

Theorem 1. *If the positive definite matrices Q and R are selected appropriately, then system (3) is asymptotically stable under the robust controller (14).*

Proof. Choose the Lyapunov function candidate as follows:

$$V = V^*(x), \tag{15}$$

which, according to (9), implies

$$\begin{aligned}
\dot{V} &= \dot{V}^*(x) \\
&= \nabla V^{*T}(x)\left(Ax + B\left(u^*(x) + \Gamma(x)\right)\right) \\
&= \nabla V^{*T}(x)\left(Ax + Bu^*(x)\right) + \nabla V^{*T}(x) B\Gamma(x) \\
&= -x^T Q x - u^{*T}(x) R u^*(x) + \nabla V^{*T}(x) B\Gamma(x).
\end{aligned} \tag{16}$$

Substituting (8) into (16) yields

$$\begin{aligned}
\dot{V} &= -x^T Q x - \frac{1}{4}\nabla V^{*T}(x) B R^{-1} B^T \nabla V^*(x) + \nabla V^{*T}(x) B\Gamma(x) \\
&\le -x^T Q x + \frac{1}{2}\Gamma^T(x)\Gamma(x) - \frac{1}{4}\nabla V^{*T}(x) B R^{-1} B^T \nabla V^*(x) \\
&\quad + \frac{1}{2}\nabla V^{*T}(x) B B^T \nabla V^*(x) \\
&\le -\left(\lambda_{\min}(Q) - \frac{1}{2}k_d^2\right)\|x\|^2 - \left(\frac{1}{4}\lambda_{\min}\left(R^{-1}\right) - \frac{1}{2}\right)\left\|B^T \nabla V^*(x)\right\|^2,
\end{aligned} \tag{17}$$

where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a matrix.

To guarantee $\dot{V} \le 0$, one should choose the parameters Q and R to satisfy the following inequalities:

$$\begin{cases}
\lambda_{\min}(Q) \ge \dfrac{1}{2}k_d^2, \\
\lambda_{\min}\left(R^{-1}\right) \ge 2.
\end{cases} \tag{18}$$

The proof is completed. □

Remark 1. By using ADP methods, one can obtain the approximate optimal control policy. However, these ADP methods are finally implemented by NNs or other universal approximators, which will bring approximation errors. In the previous works, NN approximation errors were rarely

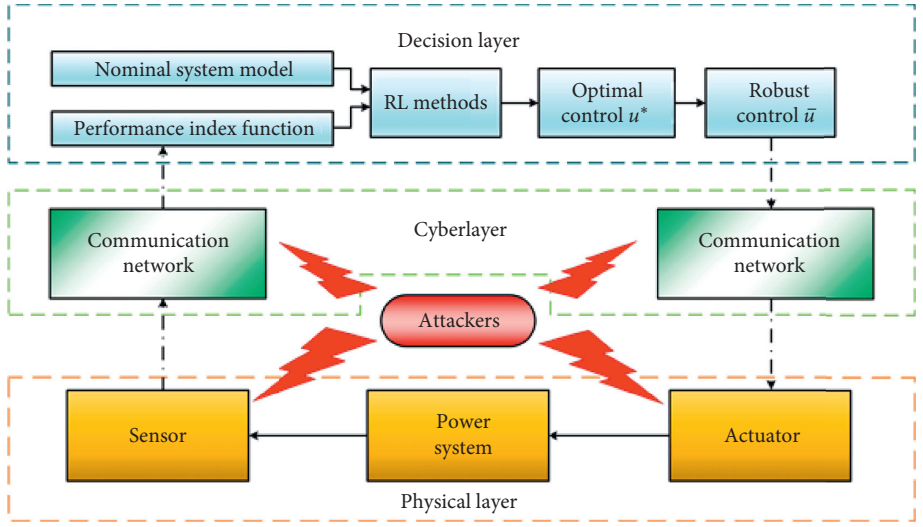


Figure 1: Block diagram of proposed robust control.

discussed. In this paper, we attempt to present the corresponding error analysis.

When NNs finish learning, NN weights will achieve convergence. Based on (11), the NN-based approximate optimal control policy, which is actually applied to the system, is expressed as

$$\hat{u}(x) = \phi_a(x)\hat{W}_a, \tag{19}$$

where $\hat{W}_a$ is the estimation of the ideal NN weight W_a. Let the approximation error be $\tilde{W}_a = \hat{W}_a - W_a$ with $\|\tilde{W}_a\| \le \tilde{W}_{am}$.

By means of (13) and (19), one gets

$$\begin{aligned}\hat{u}(x) &= \phi_a(x)W_a + \phi_a(x)\hat{W}_a - \phi_a(x)W_a \\ &= u^*(x) + \phi_a(x)\tilde{W}_a.\end{aligned} \tag{20}$$

Corollary 1. *If the positive definite matrices Q and R are selected appropriately, then the system (3) is asymptotically stable under the NN-based approximate optimal controller (19) as the NN weight approximation error $\tilde{W}_a$ goes to zero.*

Proof. Utilizing the Lyapunov function candidate (15) yields

$$\begin{aligned}\dot{V} &= \nabla V^{*T}(x)(Ax + B(\hat{u} \\ &= \nabla V^{*T}(x)\left(Ax + B\left(u^*(x) + \Gamma(x)\right)\right) + \nabla V^{*T}(x)B\phi_a(x)\tilde{W}_a.\end{aligned} \tag{21}$$

Through the result of (17), equation (21) becomes

$$\begin{aligned}\dot{V} &\le -\left(\lambda_{\min}(Q) - \frac{1}{2}k_d^2\right)\|x\|^2 - \left(\frac{1}{4}\lambda_{\min}\left(R^{-1}\right) - \frac{1}{2}\right)\left\|B^T\nabla V^*(x)\right\|^2 + \frac{1}{2}\left\|B^T\nabla V^*(x)\right\|^2 \\ &+ \frac{1}{2}\left\|\phi_a(x)\tilde{W}_a\right\|^2 \le -\left(\lambda_{\min}(Q) - \frac{1}{2}k_d^2 - \frac{1}{2}k_a^2\tilde{W}_{am}^2\right)\|x\|^2 - \left(\frac{1}{4}\lambda_{\min}\left(R^{-1}\right) - 1\right)\left\|B^T\nabla V^*(x)\right\|^2,\end{aligned} \tag{22}$$

where $\|\phi_a(x)\| \le k_a\|x\|$. If $\dot{V} \le 0$, the following condition should be satisfied:

$$\begin{cases} \lambda_{\min}(Q) \ge \frac{1}{2}k_d^2 + \frac{1}{2}k_a^2\widetilde{W}_{am}^2, \\ \lambda_{\min}\left(R^{-1}\right) \ge 4. \end{cases} \tag{23}$$

It can be observed that if the NN weight approximation error $\widetilde{W}_a$ goes to zero or is small enough, condition (23) can be easily realized with the chosen parameters. That is, the NN-based approximate optimal control policy can stabilize system (3). □

5. Robust Control Strategies for Sensor Attacks

In this section, the proposed robust control schemes are modified and extended to deal with sensor attacks [31].

5.1. Extension to Nonlinear Sensor Attacks. Consider the system (4) with nonlinear sensor attacks:

$$\dot{x} = Ax + B\overline{u} + \Gamma(x). \tag{24}$$

The robust controller for (24) is designed the same as (14), i.e., $\overline{u}(x) = u^*(x)$.

Corollary 2. *If the matrix Q is selected appropriately, then the system (24) is asymptotically stable under the robust controller $\overline{u}(x)$.*

Proof. Choose the Lyapunov function candidate as (15). Then, one attains

$$\begin{aligned} \dot{V} &= \nabla V^{*T}(x)\left(Ax + Bu^*(x) + \Gamma(x)\right) \\ &= -x^TQx - u^{*T}(x)Ru^*(x) + \nabla V^{*T}(x)\Gamma(x) \\ &\le -x^TQx - u^{*T}(x)Ru^*(x) + \frac{1}{2}\Gamma^T(x)\Gamma(x) + \frac{1}{2}\nabla V^{*T}(x)\nabla V^*(x) \\ &\le -\left(\lambda_{\min}(Q) - \frac{1}{2}k_d^2 - \frac{1}{2}k_v^2\right)\|x\|^2 - u^{*T}(x)Ru^*(x), \end{aligned} \tag{25}$$

where $\|\nabla V^*(x)\| \le k_v\|x\|$.

To guarantee $\dot{V} \le 0$, select the matrix Q to satisfy the following inequality:

$$\lambda_{\min}(Q) \ge \frac{1}{2}k_d^2 + \frac{1}{2}k_v^2. \tag{26}$$

The proof is completed.

Note that the robust load frequency control problem is a special case of sensor attacks.

Let (24) be rewritten as

$$\dot{x} = Ax + B\overline{u} + \Delta P_d(t), \tag{27}$$

where $\Delta P_d(t)$ denotes the disturbance caused by the load demand change with $\|\Delta P_d(t)\| \le \Delta P_{dm}$. □

Corollary 3. *If the matrix Q is selected appropriately, then the system states of (27) are uniformly ultimately bounded under the robust controller $\overline{u}(x)$.*

According to (15), one gets

$$\begin{aligned} \dot{V} &= \nabla V^{*T}(x)\left(Ax + Bu^*(x) + \Delta P_d(t)\right) \\ &\le -\left(\lambda_{\min}(Q) - \frac{1}{2}k_v^2\right)\|x\|^2 - u^{*T}(x)Ru^*(x) + \frac{1}{2}\Delta P_{dm}^2. \end{aligned} \tag{28}$$

Let $z = \begin{bmatrix} x \\ u^*(x) \end{bmatrix}$, $\Theta = \begin{bmatrix} (\lambda_{\min}(Q) - (1/2)k_v^2)I_x & 0 \\ 0 & R \end{bmatrix}$ with the identity matrix I_x, and $c_m = (1/2)\Delta P_{dm}^2$. If there exists a matrix Q which guarantees Θ to be positive definite, then (28) can be rewritten as

$$\begin{aligned} \dot{V} &\le -\begin{bmatrix} x \\ u^*(x) \end{bmatrix}^T \begin{bmatrix} \left(\lambda_{\min}(Q) - \frac{1}{2}k_v^2\right)I_x & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} x \\ u^*(x) \end{bmatrix} + \frac{1}{2}\Delta P_{dm}^2 \\ &\le -\lambda_{\min}(\Theta)\|z\|^2 + c_m. \end{aligned} \tag{29}$$

From (29), it can be observed that $\dot{V} \le 0$ if $\|z\| \ge \sqrt{(c_m/\lambda_{\min}(\Theta))}$. That is, the system states are uniformly ultimately bounded according to the Lyapunov extension theorem [27, 32, 33].

5.2. Extension to Constant Sensor Attacks. Consider the system (2) with constant sensor attacks:

$$\dot{x} = Ax + B\overline{u} - \Xi, \tag{30}$$

where $\dot{\Xi} = 0$.

Let $\overline{u} = u^*$ and add an attack compensator $\widehat{\Xi}$ to (30). Then, (30) becomes

$$\dot{x} = Ax + Bu^* + \widehat{\Xi} - \Xi, \tag{31}$$

where $\dot{\widehat{\Xi}} = \beta(-x^T - \nabla V^{*T}(x))^T$.

Theorem 2. *If the positive definite matrices Q and R are selected appropriately, then the system (31) is asymptotically stable under the optimal controller and the attack compensator.*

Proof. Construct a Lyapunov function candidate as follows:

$$V = \frac{1}{2}x^Tx + V^*(x) + \frac{1}{2\beta}\widetilde{\Xi}^T\widetilde{\Xi}, \tag{32}$$

where $\widetilde{\Xi} = \Xi - \widehat{\Xi}$. Then, one has

$$\begin{aligned}\dot{V} &= x^T\left(Ax + Bu^* - \widetilde{\Xi}\right) + \nabla V^{*T}(x)\left(Ax + Bu^* - \widetilde{\Xi}\right) - \frac{1}{\beta}\dot{\widetilde{\Xi}}^T\widetilde{\Xi} \\ &= x^TAx + x^TBu^* - \left(x^T + \nabla V^{*T}(x)\right)\widetilde{\Xi} + \nabla V^{*T}(x)\left(Ax + Bu^*\right) - \frac{1}{\beta}\beta\left(-x^T - \nabla V^{*T}(x)\right)\widetilde{\Xi}.\end{aligned} \tag{33}$$

After some mathematical derivation, equation (33) becomes

$$\begin{aligned}\dot{V} &= x^TAx + x^TBu^* - x^TQx - u^{*T}Ru^* \\ &\le A_m\|x\|^2 + \frac{1}{2}\|x\|^2 + \frac{1}{2}B_m^2\left\|u^*\right\|^2 - \lambda_{\min}(Q)\|x\|^2 - \lambda_{\min}(R)\left\|u^*\right\|^2 \\ &= -\left(\lambda_{\min}(Q) - A_m - \frac{1}{2}\right)\|x\|^2 - \left(\lambda_{\min}(R) - \frac{1}{2}B_m^2\right)\left\|u^*\right\|^2,\end{aligned} \tag{34}$$

where $\|A\| \le A_m$ and $\|B\| \le B_m$.

To ensure $\dot{V} \le 0$, one should set the parameters Q and R to satisfy the following inequalities:

$$\begin{cases}\lambda_{\min}(Q) \ge A_m + \dfrac{1}{2}, \\ \lambda_{\min}(R) \ge \dfrac{1}{2}B_m^2.\end{cases} \tag{35}$$

This completes the proof.

When NNs finish learning, the approximate optimal value function can be acquired:

$$\widehat{V}(x) = \phi_c^T(x)\widehat{W}_c, \tag{36}$$

where $\widehat{W}_c$ is the estimation of the ideal NN weight W_c. Let the approximation error be $\widetilde{W}_c = \widehat{W}_c - W_c$ with $\|\widetilde{W}_c\| \le \widetilde{W}_{cm}$.

Based on (31) and (36), the NN-based robust control scheme should be designed by

$$\begin{aligned}\dot{x} &= Ax + B\widehat{u} + \widehat{\Xi} - \Xi \\ &= Ax + Bu^*(x) + B\phi_a(x)\widetilde{W}_a + \widehat{\Xi} - \Xi,\end{aligned} \tag{37}$$

where

$$\begin{aligned}\dot{\widehat{\Xi}} &= \beta\left(-x^T - \nabla\widehat{V}^T(x)\right)^T \\ &= \beta\left(-x^T - \nabla V^{*T}(x) - \widetilde{W}_c^T\nabla\phi_c(x)\right)^T.\end{aligned} \tag{38}$$

□

Corollary 4. *If the positive definite matrices Q and R are selected appropriately, then the NN-based robust control scheme can stabilize the system (37) as the NN weight approximation errors $\widetilde{W}_a$ and $\widetilde{W}_c$ go to zero.*

Proof. Employing the Lyapunov function candidate (32) yields

$$\begin{aligned}\dot{V} &= x^T\left(Ax + B\left(u^* + \phi_a(x)\widetilde{W}_a\right) - \widetilde{\Xi}\right) + \nabla V^{*T}(x)\left(Ax + B\left(u^* + \phi_a(x)\widetilde{W}_a\right) - \widetilde{\Xi}\right) \\ &\quad - \frac{1}{\beta}\beta\left(-x^T - \nabla V^{*T}(x) - \widetilde{W}_c^T\nabla\phi_c(x)\right)\widetilde{\Xi} \\ &= x^TAx + x^TBu^* + \nabla V^{*T}(x)\left(Ax + Bu^*\right) + \left(x + \nabla V^*(x)\right)^TB\phi_a(x)\widetilde{W}_a + \widetilde{W}_c^T\nabla\phi_c(x)\widetilde{\Xi} \\ &\le -\left(\lambda_{\min}(Q) - A_m - \frac{1}{2}\right)\|x\|^2 - \left(\lambda_{\min}(R) - \frac{1}{2}B_m^2\right)\left\|u^*\right\|^2 \\ &\quad + \left(x + \nabla V^*(x)\right)^TB\phi_a(x)\widetilde{W}_a + \widetilde{W}_c^T\nabla\phi_c(x)\widetilde{\Xi}.\end{aligned} \tag{39}$$

From (39), in the limit as the NN weight approximation errors go to zero, i.e., $\widetilde{W}_a \longrightarrow 0$ and $\widetilde{W}_c \longrightarrow 0$, one can easily set the parameters Q and R to guarantee $\dot{V} \le 0$. If the NN approximation errors are not small enough or NNs fail to approximate optimal values, the result may not be asymptotically stable and the robust control scheme may be invalid. Therefore, the design of ADP learning algorithm is the key point. □

Table 1: Values of system parameters.

Parameters	T_g	T_t	T_p	k_s	k_p
Values	1	1	1	0.2	2

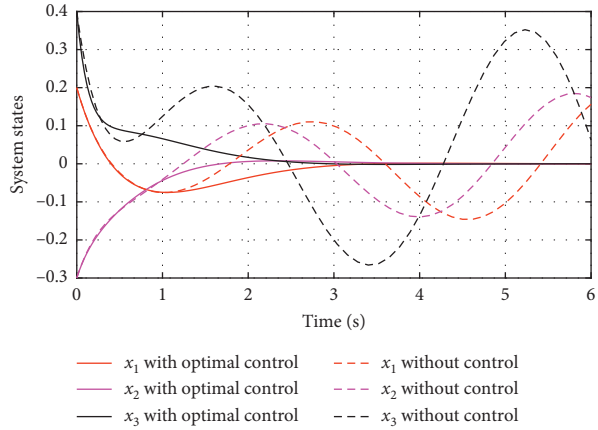


Figure 2: System states under actuator attacks.

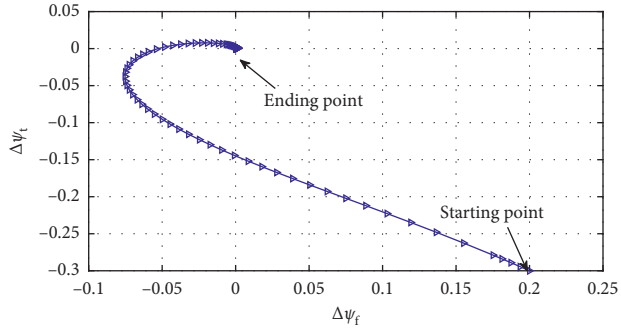


Figure 3: 2D plot of $\Delta\psi_f$ and $\Delta\psi_t$.

6. Simulation Example

In this section, to verify the proposed robust control strategy, two simulation examples of power systems are presented for two different types of attacks, respectively.

6.1. Design against Actuator Attacks. In this case, the actuator attack affecting the controller is considered in power system. The values of system parameters for this simulation are given in Table 1. Then, we can obtain system matrices $A = \begin{bmatrix} -1 & 2 & 0 \\ 0 & -1 & 1 \\ -5 & 0 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Let the initial system state values be $x(0) = [0.2; -0.3; 0.4]$. When we insert the actuator attacks $\Lambda = \Gamma(x)$ into the system, the states of power systems become unstable, which is shown in Figure 2.

In this case, the parameters are selected as $Q = 10I_3$ and $R = 0.2$, respectively. One can attain the optimal control $u^*(x)$ via Matlab command CARE or other RL methods. By using the robust controller (14), the system states under actuator attacks can be stabilized within 6 seconds in Figure 2. The 2D plot of state convergence trajectory is given in Figure 3, which indicates the nice performance of our control design.

Table 2: Values of system parameters.

Parameters	T_g	T_t	T_p	k_s	k_p
Values	0.5	0.5	0.5	1	1

6.2. Design against Constant Sensor Attacks. In this case, the designed controller is proved by numerical simulation results that it can effectively resist the sensor attacks. The values of system parameters for this simulation are given in Table 2.

Then, we can obtain system matrices $A = \begin{bmatrix} -2 & 2 & 0 \\ 0 & -2 & 2 \\ -2 & 0 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$. The controller parameters are selected as $Q = 8I_3$ and $R = 5$. Let the initial system state values be

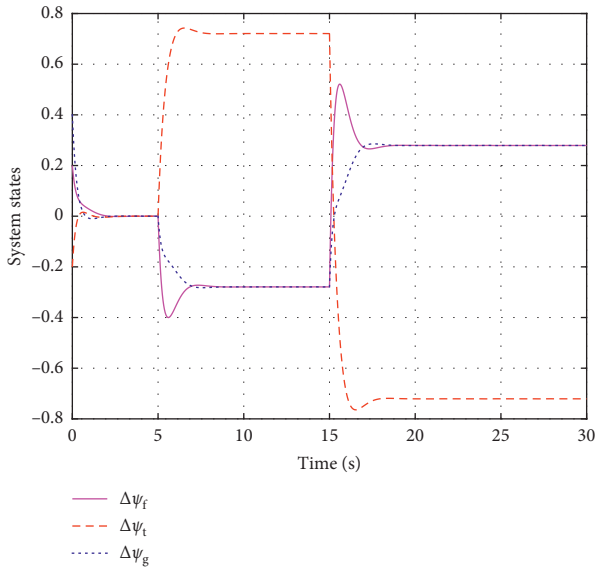


Figure 4: System states without the attack compensator.

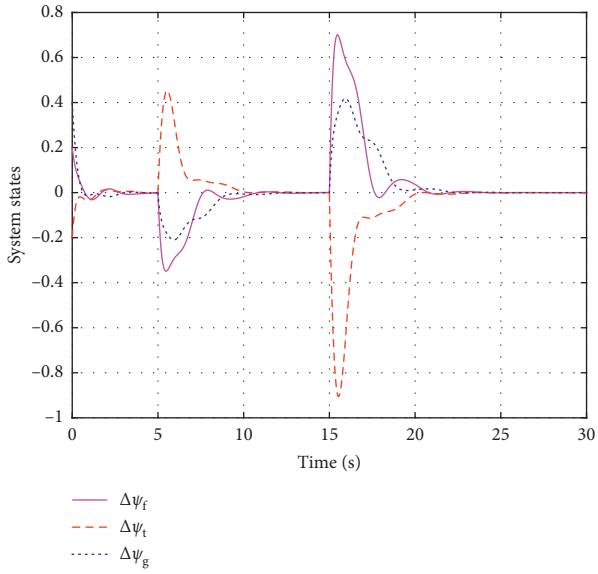


Figure 5: System states with the attack compensator.

$x(0) = [0.2; -0.2; 0.4]$ and constant sensor attack be $\Xi = \begin{cases} [0,0,0]^T & 0s \le t \le 5s \\ [2,-2,1]^T & 5s < t \le 15s \\ [-2,2,-1]^T & 15s < t \le 30s \end{cases}$.

First, we present the simulation result without the attack compensator in Figure 4, which implies the system states affected by constant attacks converge to unexpected values. Then, we employ the attack compensator-based robust control scheme, and simulation results are obtained in Figure 5, which shows the proposed scheme can quickly stabilize the system after the attacks occur. Figure 6 displays the dynamics of the attack compensator. Compared with the given attack, the compensator can estimate the attack value in a short time, which indicates the compensator can successfully get rid of the impact caused by the constant sensor attacks.

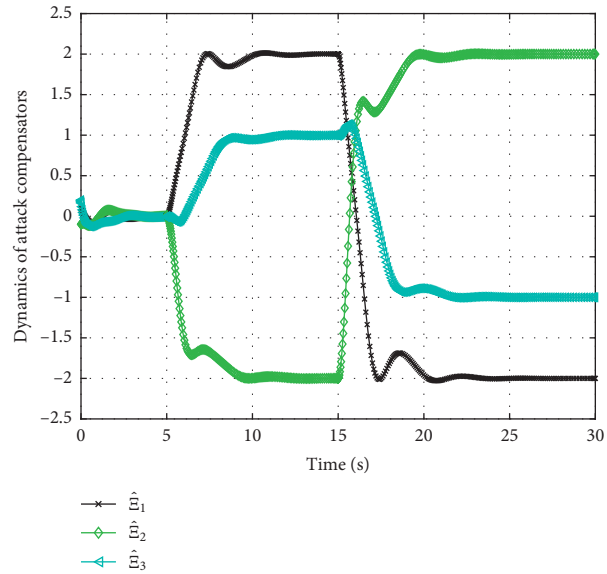


Figure 6: Dynamics of the attack compensator.

7. Conclusions

This paper has integrated optimal control theory, RL, and NNs to address the robust control issues of a benchmark power system. The optimal control theory for nominal systems and state-of-the-art RL methods along with the NN implementations have been reviewed. Multiple types of attacks in power systems, such as actuator attacks, nonlinear sensor attacks, and constant sensor attacks, are discussed. Then, several robust control schemes have been designed for different types of attacks, respectively. The control parameters have been derived through the Lyapunov stability theory. Furthermore, the stability analysis with the NN approximation error, which is rarely discussed in the previous works, has been presented in this paper. Simulation results have demonstrated the effectiveness of our proposed schemes.

Data Availability

Data are available upon request to the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Science and Technology Foundation of SGCC (SGLNDK00DWJS1900036), the Liaoning Revitalization Talents Program (XLYC1907138), the Doctoral Scientific Research Foundation of Liaoning Province (2020-BS-181), the Natural Science Foundation of Liaoning Province (2019-MS-239), the Key R&D Program of

Liaoning Province (2020JH2/10300101), the Technology Innovation Talent Fund of Shenyang (RC190360), and the Science and Technology Project of State Grid Liaoning Electric Power Company Limited (SGLNSY00HLJS2002775).

References

[1] X. Jin, W. M. Haddad, and T. Yucelen, "An adaptive control architecture for mitigating sensor and actuator attacks in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 6058–6064, 2017.

[2] G. Volkan, P. Steffen, G. Tony, and V. Frank, "A survey on concepts, applications, and challenges in cyber-physical systems," *KSII Transactions on Internet and Information Systems*, vol. 8, no. 12, pp. 4242–4268, 2014.

[3] R. Wang, Q. Sun, D. Ma, and X. Hu, "Line impedance cooperative stability region identification method for grid-tied inverters under weak grids," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 2856–2866, 2020.

[4] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 Ukraine blackout: implications for false data injection attacks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3317-3318, 2017.

[5] G. Liang, S. R. Weller, F. Luo, J. Zhao, and Z. Y. Dong, "Distributed blockchain-based data protection framework for modern power systems against cyber attacks," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3162–3173, 2019.

[6] R. Wang, Q. Sun, D. Ma, D. Qin, Y. Gui, and P. Wang, "Line inductance stability operation domain assessment for weak grids with multiple constant power loads," *IEEE Transactions on Energy Conversion*, 2020.

[7] Y. Yuan, H. Yuan, L. Guo, H. Yang, and S. Sun, "Resilient control of networked control system under DoS attacks: a unified game approach," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 5, pp. 1786–1794, 2016.

[8] Y. Yuan, F. Sun, and H. Liu, "Resilient control of cyber-physical systems against intelligent attacker: a hierarchal stackelberg game approach," *International Journal of Systems Science*, vol. 47, no. 9, pp. 2067–2077, 2015.

[9] D. Seo, H. Lee, A. Perrig, and APFS, "APFS: adaptive probabilistic filter scheduling against distributed denial-of-service attacks," *Computers & Security*, vol. 39, pp. 366–385, 2013.

[10] H. Zhang, P. Cheng P, L. Shi, and J. Chen, "Optimal DoS attack scheduling in wireless networked control system," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 2, pp. 843–852, 2016.

[11] A. Chattopadhyay and U. Mitra, "Security against false data-injection attack in cyber-physical systems," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 2, pp. 1015–1027, 2020.

[12] Y. Guan and X. Ge, "Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks," *IEEE Transactions on Signal and Information Processing Over Networks*, vol. 4, no. 1, pp. 48–59, 2018.

[13] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *Proceedings of the 2012 IEEE Global Communications Conference*, pp. 3153–3158, Anaheim, CA, USA, December 2012.

[14] A. Anwar, A. N. Mahmood, and Z. Tari, "Identification of vulnerable node clusters against false data injection attack in an AMI based smart grid," *Information Systems*, vol. 53, pp. 201–212, 2015.

[15] J. Park, R. Ivanov, J. Weimer, M. Pajic, and I. Lee, "Sensor attack detection in the presence of transient faults," in *Proceedings of the 2015 ACM/IEEE 6th International Conference on Cyber-Physical Systems*, pp. 1–10, Seattle, WA, USA, April 2015.

[16] P. M. Lima, L. K. Carvalho, and M. V. Moreira, "Detectable and undetectable network attack security of cyber-physical systems," *IFAC-PapersOnLine*, vol. 51, no. 7, pp. 179–185, 2018.

[17] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal denial-of-service attack scheduling with energy constraint," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 3023–3028, 2015.

[18] L. Che, X. Liu, Z. Li, and Y. Wen, "False data injection attacks induced sequential outages in power systems," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1513–1523, 2019.

[19] L. K. Carvalho, Y.-C. Wu, R. Kwong, and S. Lafortune, "Detection and mitigation of classes of attacks in supervisory control systems," *Automatica*, vol. 97, pp. 121–133, 2018.

[20] K. Pan, A. Teixeira, M. Cvetkovic, and P. Palensky, "Cyber risk analysis of combined data attacks against power system state estimation," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3044–3056, 2019.

[21] R. Wang, Q. Sun, D. Ma, and Z. Liu, "The small-signal stability analysis of the droop-controlled converter in electromagnetic timescale," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1459–1469, 2019.

[22] X. Huang and J. Dong, "Reliable control policy of cyber-physical systems against a class of frequency-constrained sensor and actuator attacks," *IEEE Transactions on Cybernetics*, vol. 48, no. 12, pp. 3432–3439, 2018.

[23] D. Wang, C. Li, D. Liu, and C. Mu, "Data-based robust optimal control of continuous-time affine nonlinear systems with matched uncertainties," *Information Sciences*, vol. 366, pp. 121–133, 2016.

[24] D. Wang, D. Liu, H. Li, and H. Ma, "Neural-network-based robust optimal control design for a class of uncertain nonlinear systems via adaptive dynamic programming," *Information Sciences*, vol. 282, pp. 167–179, 2014.

[25] X. Yang, D. Liu, B. Luo, and C. Li, "Data-based robust adaptive control for a class of unknown nonlinear constrained-input systems via integral reinforcement learning," *Information Sciences*, vol. 369, pp. 731–747, 2016.

[26] B. Zhao, D. Liu, and Y. Li, "Observer based adaptive dynamic programming for fault tolerant control of a class of nonlinear systems," *Information Sciences*, vol. 384, pp. 21–33, 2017.

[27] K. G. Vamvoudakis and F. L. Lewis, "Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, 2010.

[28] D. Wang, H. He, C. Mu, and D. Liu, "Intelligent critic control with disturbance attenuation for affine dynamics including an application to a microgrid system," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 4935–4944, 2017.

[29] D. Wang, H. He, X. Zhong, and D. Liu, "Event-driven nonlinear discounted optimal regulation involving a power system application," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 10, pp. 8177–8186, 2017.

[30] X. Zhong and H. He, "An event-triggered ADP control approach for continuous-time system with unknown internal states," *IEEE Transactions on Cybernetics*, vol. 47, no. 3, pp. 683–694, 2017.

[31] X. Yang and H. He, "Self-learning robust optimal control for continuous-time nonlinear systems with mismatched disturbances," *Neural Networks*, vol. 99, pp. 19–30, 2018.

[32] Q. Wei, R. Song, and P. Yan, "Data-driven zero-sum neuro-optimal control for a class of continuous-time unknown nonlinear systems with disturbance using ADP," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 2, pp. 444–458, 2016.

[33] D. Zhao, Q. Zhang, D. Wang, and Y. Zhu, "Experience replay for optimal control of nonzero-sum game systems with unknown dynamics," *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 854–865, 2016.

Hindawi
Complexity
Volume 2021, Article ID 5536294, 11 pages
https://doi.org/10.1155/2021/5536294

Research Article

Static Stability Analysis Based on Probabilistic Power Flow Calculation considering P2G Technology

Jia Yao, Yujia Huang, and Jingwei Hu

College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China

Correspondence should be addressed to Jia Yao; 1970543@stu.neu.edu.cn

Received 17 January 2021; Revised 26 January 2021; Accepted 2 March 2021; Published 11 March 2021

Academic Editor: Dan Selisteanu

At present, integrated energy systems have received extensive attention, but there is no basic framework for stability analysis of coupled systems. The injection of a large amount of renewable energy also has a great impact on the stability of the system. This paper focuses on how to analyze the static stability of the coupling system with uncertainty, which mainly considers the uncertainty of wind power generation and photovoltaic power generation and also considers the influence of P2G technology on the whole system. Firstly, this paper analyzes the principles of wind power generation and photovoltaic power generation and constructs the probability model of renewable energy power generation power. Then, the three-point estimation method is used to process the data, and the probability distribution of the unknown quantity is obtained by probabilistic power flow analysis. Finally, the probability distribution of each eigenvalue is obtained by analyzing the sensitivity of the characteristic roots to the voltage. Thus, the static stability of the system is judged. The applicability of proposed methodology is demonstrated by analyzing an integrated IEEE 14-bus power system and a Belgian 20-node gas system in this paper.

1. Introduction

With the in-depth study of renewable energy [1, 2], wind power generation and photovoltaic power generation are connected to the grid in large numbers. But, the intermittent and random nature of renewable energy will have a great impact on the power grid.

The peak period of renewable energy is probably the low valley period of power grid load resulting in great waste. Considering that the reaction is unlikely to be complete, P2G technology will eventually produce a gas mixture of hydrogen and methane [3–6]. At present, many studies show that existing natural gas pipelines can be used to accept mixed gases generated by P2G technology [7]. Moreover, the mixed gas can be fed into the gas network to realize the bidirectional conversion between gas and electricity, which increases the coupling between the gas network and the power network. In this way, cross-regional and large-scale interconnection between electricity and gas can be realized, and the advantages of optimal energy allocation can also be realized [8, 9].

And, randomness is the uncertainty of exponential value. The generation power of renewable energy depends on the amount of energy at that time, and the amount of renewable energy is affected by many natural factors and is not controllable, which brings uncertain data to the system. This makes it difficult to judge the state of the system. It is very important whether the system can run stably [10, 11]. The common way to obtain the running state of the system is the traditional power flow calculation, but the traditional power flow can not deal with the uncertain data brought by renewable energy, so it is necessary to use the uncertain power flow calculation to obtain the running state of the system.

At present, there are three methods for calculating probabilistic power flow: fuzzy mathematics method [12, 13], probabilistic power flow method [14], and interval analysis power flow method. The basic assumption of fuzzy distribution power flow is that the power demand boundary of the system bus is uncertain and obeys a certain probability distribution. Its solution set is some set of uncertain boundaries. The basic assumption of probabilistic power flow and interval power flow is that the power demand of the system bus is uncertain in a given interval. Among them, the basic mathematical theory of probabilistic power flow is probability theory. It represents uncertain data as a

probability distribution and calculates its mean and variance. The uncertainty of known quantities is dealt by the Monte Carlo simulation method, semi-invariant method, point estimation method, and other data processing methods. Then, the probability distribution of the unknown quantity can be calculated by the power flow. It takes into account the random factors in system operation and can better represent the characteristics of power system operation than traditional power flow calculation, so as to facilitate us to find out the potential dangers and weak links of system operation.

Many papers have studied the probabilistic power flow calculation of wind power generation and photovoltaic power generation. In [15], the definition of probabilistic power flow calculation is proposed, and the most basic solution method is given. In [16], using the knowledge of probability theory, the probability distribution of the known quantity is analyzed, and the probability distribution of the unsolved quantity is finally obtained. The Markov chain quasi-Monte Carlo sampling method is adopted, and the correlation of power generation efficiency of multiple wind farms is considered to obtain the final probabilistic optimal power flow, in [17]. In consideration of the incomplete acceptance of wind power generation, in [18], a probabilistic power flow calculation method is propose based on the principle of proportional distribution and Latin hypercube sampling, which ensures the calculation accuracy and greatly improves the calculation speed.

The state of the system is often obtained by power flow calculation to judge the static stability of the system. Static stability refers to the ability of a system to operate stably after a small disturbance, which can be used to find the weak points of the system. In order to ensure the safe and reliable operation of the system, it is necessary to analyze the static stability of the system [19–21]. The randomness and intermittency of renewable energy will bring great fluctuation to the system and affect the static stability of the system. However, probabilistic power flow calculation is different from traditional power flow calculation. The result of uncertain power flow calculation is the probability distribution of system running state, not a set of fixed running values. So, this causes some difficulties to the static stability analysis.

Not only the power grid has static stability but also the gas network has static stability. Because the gas transported in the gas network is flammable and explosive, the stability of the gas network is also very important. But, P2G technology will use extra power to electrolyze water to get combustible gas and put the resulting mixture into the gas network. It will lead to changes in gas characteristics specific gravity (SG) and gross calorific value (GCV) and influence the operation of the gas grid. If a more accurate stability analysis of the integrated power-gas System is required, we should further the model of the gas network with gas mixture and calculate the power flow according to the new model. There are several papers on integrated power-gas systems. Yang et al. [22] modeled integrated power-gas systems, taking into account the uncertainty of pipeline parameters. Considering the different response time of gas network and power network, the optimal operation strategy is studied by power flow calculation, in [23].

To the best of authors' knowledge, there is little consideration that the stability of a system with uncertain data, and no consideration that the uncertainty of renewable energy and the consumption of renewable energy by P2G technology's together. Of course, there also is no consideration that the static stability of integrated power-gas systems static with wind power generation and photovoltaic power generation. To solve these problems, the main contributions of this paper can be concluded as follows:

(1) An probabilistic power flow calculation method considering P2G technology is proposed. It not only considers the impact of the uncertainty of renewable energy on the system but also considers the impact of the gas mixture produced by P2G on the gas network. This makes the static stability analysis of the system more practical and more accurate.

(2) Based on probabilistic power flow calculation, the probability distribution of the power system static stability probability distribution of power system is analyzed. Considering the influence of uncertain data on power system, the confidence level of system stability can be obtained by analyzing the probability distribution of the unknown quantity. The potential dangers and weak links of power system operation can also be judged through the data.

(3) Not only the stability of electric power network but also the stability of gas network is considered. The uncertainty of renewable energy is analyzed, and its influence on the stability of gas network is obtained. In this way, it can effectively judge whether there is pipe blocking phenomenon in the gas network, so as to prevent accidents to some extent.

This paper is described in detail as follows. Treatment of uncertain data on wind and photovoltaic power generation is described in Section 2. An uncertain power flow calculation method considering P2G technology is described in Section 3. Static stability analysis based on uncertain power flow calculation is described in Section 4. Finally, the case studies are carried out on the integrated IEEE 14-bus power system and a Belgian 20-node gas system in Section 5, and conclusions are given in Section 6.

2. Data Processing

2.1. Wind Power Generation. The height of the weather station wind meter may not be the same as the height of the fan wheel shaft, so it is necessary to convert the wind speed measured by the wind meter into the wind speed at the fan wheel shaft height when calculating the wind power generated. The relationship between them can be expressed as

$$v = v_0\left(\frac{h}{h_0}\right)^a, \tag{1}$$

where v is the wind speed at fan shaft height, h is the height of fan shaft, v_0 is the wind speed as measured by a wind gauge, h_0 is the height of the wind gauge, and a is the roughness of the ground, which is usually 1/7 in the open

space. Some studies have shown that the average hourly wind speed in a year conforms to the Weibull distribution [24], but the study, which looked at 24 hours a day, suggests that turbulence, which varies in wind speed over short periods of time, cannot be ignored [25]. The intensity of turbulence can be expressed as

$$I = \frac{\sigma_v}{\mu_v}, \tag{2}$$

where μ_v is the mean of wind speed and σ_v is the standard deviation of the wind speed. Considering the influence of turbulence on wind speed, wind speed distribution can be expressed as a normal distribution. So, the wind speed probability density can be expressed as

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v} e^{-(v-\mu_v)^2/2\sigma_v^2}. \tag{3}$$

The relationship between the output power of wind power generation and wind speed can be expressed as

$$P_w = \begin{cases} 0, & v < v_{\text{in}}, \\ \dfrac{v - v_{\text{in}}}{v_N - v_{\text{in}}} P_N, & v_{\text{in}} \le v \le v_N, \\ P_N, & v_N \le v \le {}_{\text{out}}, \\ 0, & v > v_{\text{out}}, \end{cases} \tag{4}$$

where P_N is the rated power of wind turbines, v_{in}, v_N, and v_{out} are cut in wind speed, rated wind speed, and cut out wind speed, respectively. The mean and the standard deviation of the output of the wind power generation can be expressed as

$$\begin{aligned} \mu_{\text{pw}} &= \int_0^{+\infty} P_w(v) f(v) \mathrm{d}v, \\ \sigma_{\text{pw}} &= \sqrt{\int_0^{+\infty} \left(P_w(v) - \mu_{\text{pw}}\right)^2 f(v) \mathrm{d}v}. \end{aligned} \tag{5}$$

2.2. Photovoltaic Power Generation. Daily exposure hours can be obtained from the data from weather stations. The total amount of daylight radiation throughout the day can be expressed as

$$\begin{aligned} H &= H_L\left(a + b\frac{s}{s_L}\right), \\ S_L &= \frac{2}{15} W_s, \\ H_L &= \tau H_0, \end{aligned} \tag{6}$$

where H_L is the intensity of solar radiation entering the Earth's atmosphere, a and b are constant coefficients, S_L is day length, W_s is hour angle, H_0 is the radiation angle of the outer layer of the Earth's atmosphere, and τ is the air transparency coefficient. The air transparency coefficient generally fluctuates between 0.4 and 0.8. Its probability density function can be expressed as

$$f(\tau) = c\frac{\tau_{\max} - \tau}{\tau_{\max}} e^{\lambda\tau}, \tag{7}$$

where $\tau_{\max}$ is the maximum air transparency coefficient and c and λ are determined by the maximum air transparency coefficient and the mean value of air transparency coefficient. They can be represented as

$$\begin{aligned} c &= \frac{\lambda^2 \tau_{\max}}{e^{\lambda\tau_{\max}} - \lambda\tau_{\max} - 1}, \\ \lambda &= \frac{2r - 17.5119e^{-1.3118r} - 1062e^{-5.0426r}}{\tau_{\max}}, \\ r &= \frac{\tau_{\max}}{\tau_{\max} - \mu_\tau}, \end{aligned} \tag{8}$$

where μ_τ is the mean of the air transparency coefficient. Therefore, if the total daily radiation amount is known, the probability density function of a day's light intensity can be expressed as

$$R(t) = \frac{1}{\sqrt{2\pi}\sigma_R} e^{-\left((t-\mu_R)^2/2\sigma_R^2\right)}. \tag{9}$$

Most of the existing studies use peak watt power to calculate the photovoltaic power generation. The daily exposure hours can be converted to peak watt hours and expressed as

$$D(t) = 0.0116H(t), \tag{10}$$

where $D(t)$ is watts peak hours and 0.0116 is the conversion factor and in h·cm^2/cal. Therefore, the output power of the photovoltaic power can be expressed as

$$P_R(t) = P_f \eta\ D(t), \tag{11}$$

where P_f and η are the peak watt power and efficiency of the photovoltaic power generation respectively.

2.3. Three-Point Estimation Method. The three-point estimation method is the Taylor series expansion of the function $Y = h(X)$ associated with the n-dimensional random variable X. The probability density of Y is obtained by estimating Y at three points using the higher order moments of X. The taking positions and probabilities of these three points can be expressed as

$$\begin{cases} X_i = \left(\mu_x + \xi_i \cdot \sigma_x\right), & i = 1, 2, 3, \\ P_i = \dfrac{(-1)^{3-i}}{\xi_i(\xi_1 - \xi_2)}, & i = 1, 2, \\ P_3 = 1 - P_1 - P_2, \end{cases} \tag{12}$$

where μ_x and σ_x are the mean value and variance of x variable, respectively, P_i is the probability of taking the ith

point position, and ξ_i is the measurement coefficient of the position of the ith point; it can be represented as

$$\begin{cases} \xi_i = \dfrac{\alpha_3}{2} + (-1)^{3-i}\sqrt{\alpha_4 - \dfrac{3\alpha_3^2}{4}}, \quad i = 1, 2, \\ \xi_3 = 0, \end{cases} \tag{13}$$

where α_i can be expressed as

$$\begin{aligned} \alpha_i &= \frac{M_i(X)}{\sigma_x^i}, \\ M_i(X) &= \int_{-\infty}^{+\infty} \left(X - \mu_x^i\right) f(x) \mathrm{d}x, \end{aligned} \tag{14}$$

where $M_i(X)$ is the center distance and $f(x)$ is the probability density function. Because the three-point estimation method is used, the center distance and lambda coefficient are only required up to the fourth order. So, we know that $\alpha_1 = 0$ and $\alpha_2 = 1$ from the above formula. After determining the position of the point, the value of the unknown quantity at each point can be obtained by the function $Y = h(X)$. According to the probability of each point, the estimated values of the center distances of Y can be expressed as

$$E\left(Y^j\right) \approx \sum_{i=1}^{3} P_i \cdot [h(X)]^i. \tag{15}$$

So, the standard deviation of Y can be expressed as

$$\sigma_Y = \sqrt{E\left(Y^2\right) - E^2(Y)}. \tag{16}$$

The uncertain data in the system are processed by the three-point estimation method, and the three pick-point positions of each uncertain node can be obtained. Then, substituting these three points into the power flow calculation, respectively, can obtain the estimated value of each point. Because one of the three pick points for each uncertain data is the mean, so if the number of the uncertain nodes in the system is N, the number of power flow calculations will be $2N-1$.

3. Probabilistic Power Flow Calculation

3.1. Electric Network. In the electric network, in addition to the line parameter resistance and reactance, each node also corresponds to four variables, which are active power, reactive power, voltage amplitude, and voltage phase angle. According to Kirchhoff's law, the iterative equation of power flow calculation in electric network can be written as

$$\begin{cases} \Delta P_i = P_i^{\mathrm{sp}} - P_i = P_i^{\mathrm{sp}} - V_i \sum_{j \in i} V_j \left(G_{ij} \cos\, \theta_{ij} + B_{ij} \sin\, \theta_{ij}\right), \\ \Delta Q_i = Q_i^{\mathrm{sp}} - Q_i = Q_i^{\mathrm{sp}} - V_i \sum_{j \in i} V_j \left(G_{ij} \cos\, \theta_{ij} + B_{ij} \sin\, \theta_{ij}\right), \end{cases} \tag{17}$$

where G is the electric conductance, B is the electric inductance, P is the active power, Q is the reactive power, V is the voltage amplitude, and θ is the voltage phase angle.

3.2. Gas Network. In the gas network, in addition to the line parameters, pipe length, pipe diameter, and friction coefficient, each node also corresponds to two variables, namely, air pressure and gas flow rate. The pressure drop function due to line parameters can be written as

$$f_{ij}^{\mathrm{GL}} = C_{ij}^{\mathrm{GL}} \mathrm{sign}\left(\pi_i, \pi_j\right) \left(\mathrm{sign}\left(\pi_i, \pi_j\right) \cdot \left(\pi_i^2 - \pi_j^2\right)\right)^{0.5}, \tag{18}$$

where C_{ij}^{GL} is the pipeline coefficient, f is the gas-flow rate, π is the gas pressure, and $\mathrm{sign}(\pi_i, \pi_j)$ is the sign function.

According to Kirchhoff's law, considering element characteristics and network topology constraints, the flow deviation equation of gas network nodes can be written as

$$\Delta f_i = f_i^{\mathrm{GD}} + f_i^{\mathrm{GG}} + f_i^{\mathrm{CHP}} + f_i^{\mathrm{GB}} + f_i^{\mathrm{GS}} + \sum_{j=1}^{N_{\mathrm{GN}}} f_{ij}^{\mathrm{GL}} + \sum_{j=1}^{N_{\mathrm{GN}}} \tau_{ij}^{\mathrm{GC}}, \tag{19}$$

where f_i^{GD} is the gas flow from gas grid load consumption, f_i^{GG} is the gas flow from gas generator, f_i^{CHP} is the gas flow from gas turbines, f_i^{GB} is the gas flow from gas fired boiler, f_i^{GS} is the gas flow from gas source, f_{ij}^{GL} is the pipeline flow, and τ_{ij}^{GC} is the flow rate consumed by gas compressors.

3.3. P2G. Because the gas mixture makes the SG and GCV in different positions of the gas network different, there is not only one variable of pressure at each gas node; but two variables of SG and GCV need to be increased. Therefore, the known quantities of relaxation nodes in the gas network are pressure, SG, and GCV, and the unknown quantities are gas flow rate, the known quantities of the load nodes in the gas network are the energy requirements, and the unknown quantities are the pressure, SG, and GCV. There will also be nonmixed air source and mixed air source in the gas network. The known quantities of nonmixed air sources are gas flow rate, SG, and GCV, and the unknown quantities are pressure. The known quantities of the mixed air source are the gas flow rate, and the unknown quantities are the pressure, SG, and GCV.

According to the knowledge of fluid mechanics, SG and GCV of each node of gas network can also be obtained by iteration. This paper mainly considers the influence of the gas mixture of natural gas, hydrogen, and methane generated by P2G on the gas network. So, SG and GCV can be expressed as

$$\begin{aligned}
\Delta \mathrm{SG}_i &= \mathrm{SG}_i\left(f_i^N + f_i^H + f_i^M + \sum_{j=1}^{N_0} \mathrm{sg}\left(f_{ji}\right)\cdot f_{ji}\right) \\
&\quad -\left(f_i^N \mathrm{SG}_N + f_i^H \mathrm{SG}_H + f_i^M \mathrm{SG}_M + \sum_{j=1}^{N_0} \mathrm{sg}\left(f_{ji}\right)\cdot\left(f_{ji}\mathrm{GCV}_j\right)\right) = 0, \quad \forall i = 1, 2, \ldots, \left(N_G - 1 - N_{\text{non-mixed}}\right), \\
\Delta \mathrm{GCV}_i &= \mathrm{GCV}_i\left(f_i^N + f_i^H + f_i^M + \sum_{j=1}^{N_0} \mathrm{sg}\left(f_{ji}\right)\cdot f_{ji}\right) \\
&\quad -\left(f_i^H \mathrm{GCV}_N + f_i^H \mathrm{GCV}_H + f_i^M \mathrm{GCV}_M + \sum_{j=1}^{N_0} \mathrm{sg}\left(f_{ji}\right)\cdot\left(f_{ji}\mathrm{GCV}_j\right)\right) = 0, \quad \forall i = 1, 2, \ldots, \left(N_G - 1 - N_{\text{non-mixed}}\right),
\end{aligned} \tag{20}$$

where SG and GCV, whose subscripts are i and j, represents the value of the variable at node i and node j, respectively, and $\mathrm{sg}(f_{ji})$ is a sign function of gas flow where $\mathrm{sg}(f_{ji}) = 1$ if $\mathrm{sg}(f_{ji}) \geq 0$ and 0 otherwise. f_i^N, f_i^H, and f_i^M represent the gas flow of natural gas, hydrogen, and methane at node i, respectively. For more information, please refer to [3].

Therefore, the Jacobian matrix of power flow calculation for the gas mixture system can be expressed as

$$J = \begin{pmatrix}
\frac{\partial \Delta P}{\partial P} & \frac{\partial \Delta P}{\partial \theta} & \frac{\partial \Delta P}{\partial V} & \frac{\partial \Delta P}{\partial \pi} & \frac{\partial \Delta P}{\partial \mathrm{SG}} & 0 \\
0 & \frac{\partial \Delta P}{\partial \theta} & \frac{\partial \Delta P}{\partial V} & 0 & 0 & 0 \\
\frac{\partial \Delta f}{\partial P} & 0 & 0 & \frac{\partial \Delta f}{\partial \pi} & \frac{\partial \Delta f}{\partial \mathrm{SG}} & \frac{\partial \Delta f}{\partial \mathrm{GCV}} \\
\frac{\partial \Delta \mathrm{SG}}{\partial P} & 0 & 0 & \frac{\partial \Delta \mathrm{SG}}{\partial \pi} & \frac{\partial \Delta \mathrm{SG}}{\partial \mathrm{SG}} & 0 \\
\frac{\partial \Delta \mathrm{GCV}}{\partial P} & 0 & 0 & \frac{\partial \Delta \mathrm{GCV}}{\partial \pi} & 0 & \frac{\partial \Delta \mathrm{GCV}}{\partial \mathrm{GCV}}
\end{pmatrix}. \tag{21}$$

4. Static Stability Analysis

4.1. Confidence Interval and Confidence Level. A confidence interval of a probability sample is an interval estimate of a population parameter of this sample. The confidence interval shows the degree to which the true value of this parameter has a certain probability of falling around the measurement result. The confidence interval gives the confidence of the measured value of the measured parameter, that is, the "certain probability". This probability is called the confidence level.

For any system, there is no unique state in which it can reliably operate. Instead, there will be a range within which the system can operate stably. The static stability analysis is to judge whether the running state of the system at that time is within this stable range. And, the uncertainty of data will cause the fluctuation of system running state, so it is impossible to directly judge whether the running state of the system can be in a stable range. In this case, whether the running state of the system is within this range becomes a probabilistic problem. To analyze the static stability of such uncertain system is to analyze the probability that its running state is within the stable range. These are confidence intervals and confidence levels in probabilistic knowledge. The confidence interval we choose is the range that the system can operate stably. The confidence level of system stability is the probability that the system operating state is within this confidence interval.

4.2. Electric Grid Stability Judgment. Existing studies have shown that the static stability of the power system can be judged by calculating the eigenvalue of the Jacobian matrix of power flow [26]. In this paper, the probability distribution of the characteristic roots of the system can be obtained by analyzing the results of the probabilistic power flow. Then, the confidence level of the stability of the whole system can be obtained by the stability criterion. First, the characteristic roots of the Jacobian matrix calculated by each power flow are obtained, and then the mean value of each eigenvalue is obtained according to the probability of taking points. It can be represented as

$$\lambda_k \approx \sum_{i=1}^{3} P_i \cdot \lambda(x). \tag{22}$$

And, the covariance matrix of the characteristic roots can be expressed as

$$C_\lambda = J_\lambda C_v J_\lambda^T, \tag{23}$$

where C_λ is the covariance matrix of the characteristic roots, J_λ is the Jacobian matrix between the characteristic roots and the voltage, and C_v is the covariance matrix of voltage. They can be represented as

$$\begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = J_\lambda \begin{bmatrix} V_1 \\ \vdots \\ V_n \end{bmatrix} C_{vab} = \operatorname{cov}(V_a, V_b), \quad a, b = 1, 2, \ldots, n,$$
$$C_{vab} = \operatorname{cov}(V_a, V_b), \quad a, b = 1, 2 \ldots, n. \tag{24}$$

The probability distribution of each node voltage can be obtained by the above formula, so as to determine how much probability the whole power network can run stably.

4.3. Gas Grid Stability Judgment. The gas transmission pressure of natural gas will affect the operating pressure of the pipeline itself. Appropriate operating pressure is an important guarantee to ensure the stable work and safety of the pipeline. When the gas transmission pressure of natural gas is too large or greater than the normal gas transmission pressure, the gas transmission pressure will have a great impact on the pipeline, increase the flow rate of the fluid, and may directly lead to the fall off or rupture of the pipeline interface. If the gas transmission pressure of the pipeline is too small and the pressure is not enough to reach the normal operating pressure, then the gas transmission pressure cannot make the pipeline carry out normal pipeline transportation.

Therefore, in the transportation of the gas network, the pressure value of each node of the gas network has the greatest influence on the safety. The influence of the fluctuation brought by renewable energy on the stability of the gas network can be judged by judging the probability distribution of pressure of each node in the gas network:

$$E_p \approx \sum_{i=1}^{3} P_i \cdot p_k,$$
$$\sigma_p = \sqrt{E\left(p^2\right) - E^2(p)}, \tag{25}$$

where E_p and σ_p are the mean value and variance of pressure.

5. Numerical Simulations

In this paper, the feasibility of this method is verified by an integrated IEEE 14-bus power system and a Belgian 20-node system, which is shown in Figure 1. For the IEEE 14-bus power system, it has two GfG units (GfG1 at E1 and GfG2 at E2), one wind turbine at E3 and one photovoltaic generator at E6 in this paper. For the Belgian 20-node gas system [27], the gas well at G1 provides traditional natural gas and serves as a slack node with the reference pressure of 56 bar. The gas well at G5 and G8 provides synthetic natural gas and its SG and GCV are known. And, the gas well at *G*2, *G*13, and *G*14 provides the mixture of synthetic natural gas and hydrogen whose SG and GCV are unknown. It is assumed that there is no gas leakage and the gases are completely mixed without chemical reaction. The pipeline parameters and gas load demands are given in [28]. The SG and GCV of different gas supply source are shown in Table 1.

5.1. Renewable Energy Power Probability Density. The data used in this paper are from a group of wind power generation with 30 wind turbines. The rated power of each wind turbines is 2.5 MW. The cut in wind speed, rated wind speed, and cut out wind speed are, respectively, 13 m/s, 7.3 m/s, and 13 m/s. The mean value of wind speed data is 5 m/s, and the standard deviation is 8 m/s. Thus, the probability density function of the output power of the wind generator can be obtained, which is shown in Figure 2.

In this paper, the photovoltaic generator is located at 104.4°*E* and 30.4°*N*. The loss coefficient is 0.85. The average temperature during the day is 24°C. It is also possible to plot the probability density of the output power of the photovoltaic generator, which is shown in Figure 3.

5.2. Static Stability Analysis of Electric System. Using the three-point estimation method to process the data, we can obtain the three-point location and probability of wind power generation and photovoltaic power generation. And then, plugging those points into the power flow equation can get five sets of results. Finally, the probability distribution of eigenvalues calculated by each power flow can be analyzed. Because the whole power system has 14 nodes, including one slack node, four PV nodes, and nine PQ nodes, there are 22 power flow equations in the power system, namely, 22 characteristic roots. There are nine pairs of conjugate characteristic roots, the 1st and 2nd, the 4th and 5th, the 7th and 8th, the 10th and 11th, the 12th and 13th, the 14th and 15th, the 16th and 17th, the 19th and 20th, and the 21st and 22nd. Since the static stability of the system depends on the real part of each characteristic root, the effect of a pair of conjugate characteristic roots on the stability of the system is the same. The probability distribution of the characteristic root real part can be obtained by analyzing the five groups of characteristic root values, which is shown in Figure 4. According to the curve in the figure, it can be found that each characteristic root has a high probability of falling on the negative axis, so the power system has a good stability.

5.3. Static Stability Analysis of Gas System. The impact of renewable energy uncertainty on SG and GCV of each node can be analyzed through the data obtained from the above power flow calculation. Finally, the mean values of SG and GCV at each node is shown in Table 2. The fluctuation trend of SG of node 14 affected by uncertain information is shown in Figure 5, and the fluctuation trend of GCV is in Figure 6.

By analyzing the pressure value of each node in the gas network, the probability distribution can be obtained, which

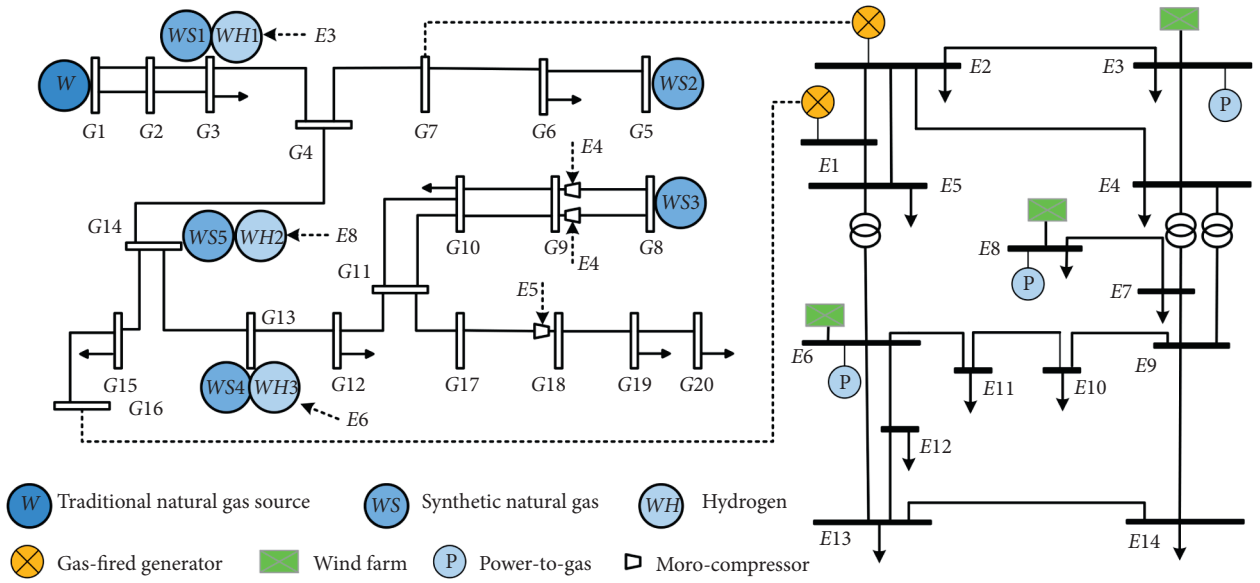


Figure 1: Integrated IEEE 14-bus system and Belgian 20-node gas system.

Table 1: SG and GCV of gas sources.

	Natural gas	Synthetic natural gas	Hydrogen
SG	0.6048	0.58	0.0696
GCV	41.04	37.4	12.75

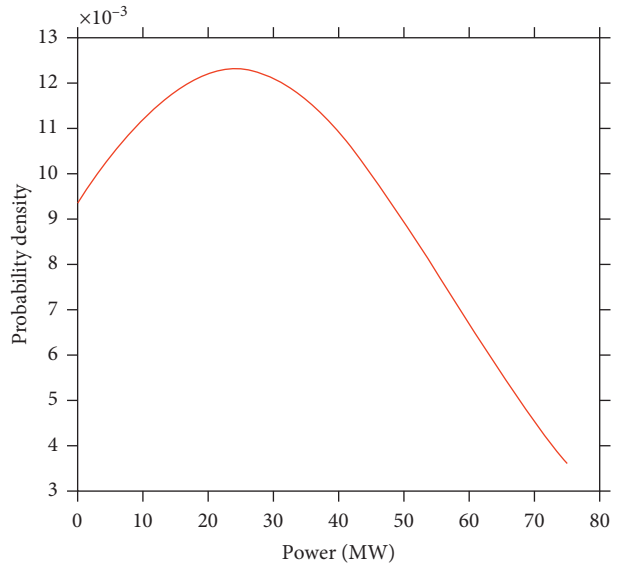


Figure 2: Probability density function of wind photovoltaic power generation.

is shown in Figure 7. It can be seen that the pressure of each node fluctuates in a certain range. The higher the ordinate, the greater the possibility that the pressure of the node falls there, the more stable the node is. It is obvious that the pressure of Node 3 is the most stable, and Figure 8 is the curve of Node 3.

Figure 9 compares Node 3, Node 19, and Node 20. It can be clearly seen that compared with Node 3, node 19, and Node 20 have great pressure changes. It can be seen that the stability of node 19 and node 20 is very poor, which is the weak link of the system. This shows that the nodes at the end of the network are very fragile. When we pay attention to the

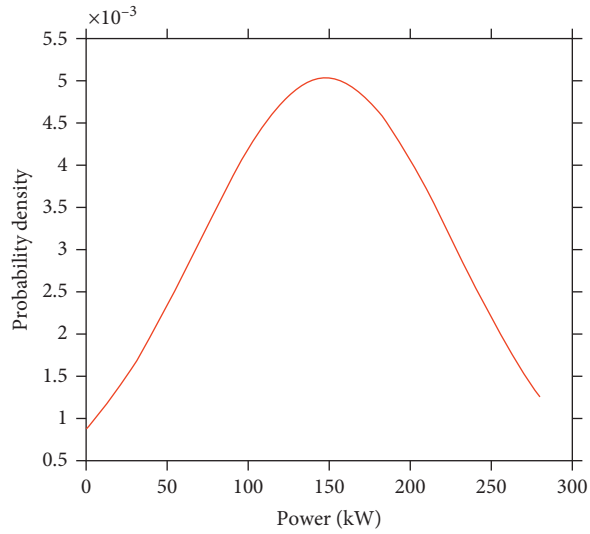


Figure 3: Probability density function of photovoltaic power generation.

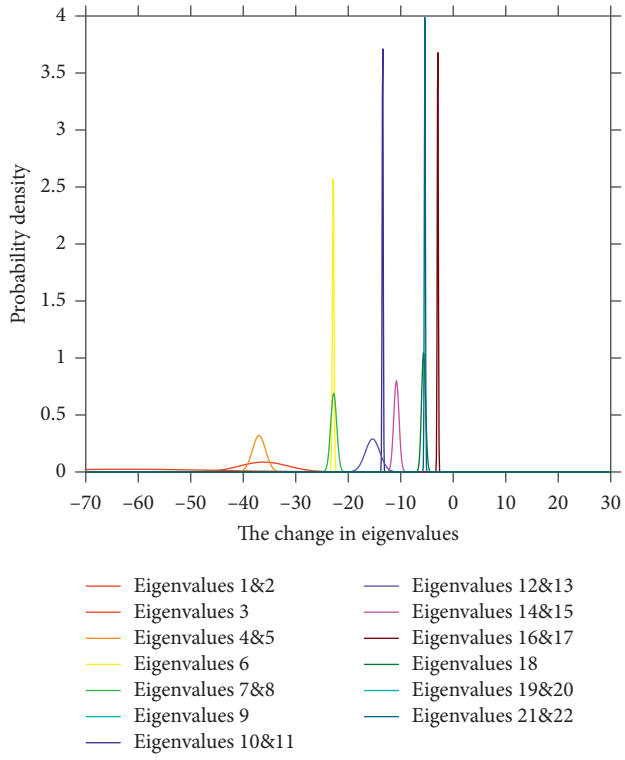


Figure 4: Probability distribution of the eigenvalue.

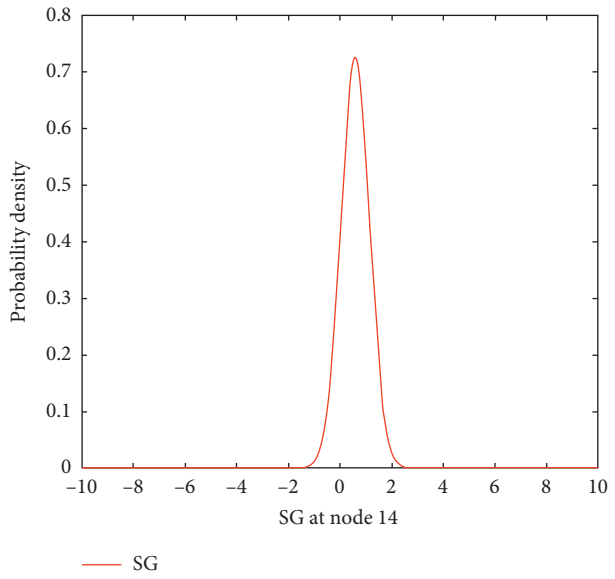


Figure 5: SG at node 14.

Table 2: SG and GCV at different nodes.

Node	SG	GCV	Node	SG	GCV
1	0.6048	41.04	11	0.58	37.4
2	0.5972	39.9413	12	0.58	37.4
3	0.5972	39.9413	13	0.58	37.398
4	0.5972	39.9413	14	0.5883	38.6338
5	0.58	37.4	15	0.5883	38.6338
6	0.5861	38.2952	16	0.5883	38.6338
7	0.5972	39.9413	17	0.58	37.4
8	0.58	37.4	18	0.58	37.4
9	0.58	37.4	19	0.58	37.4
10	0.58	37.4	20	0.58	37.4

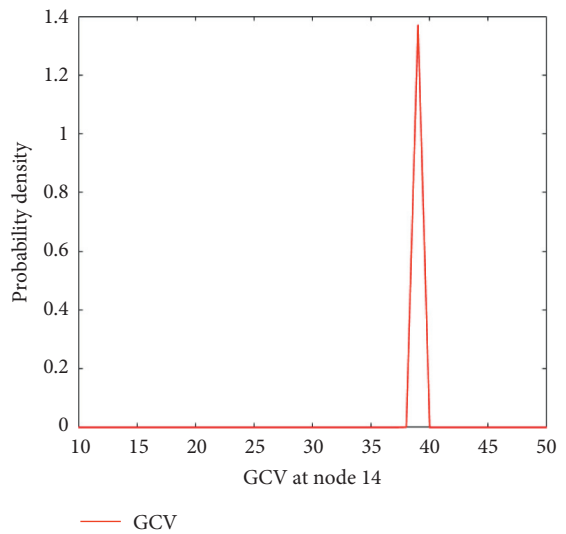


Figure 6: GCV at node 14.

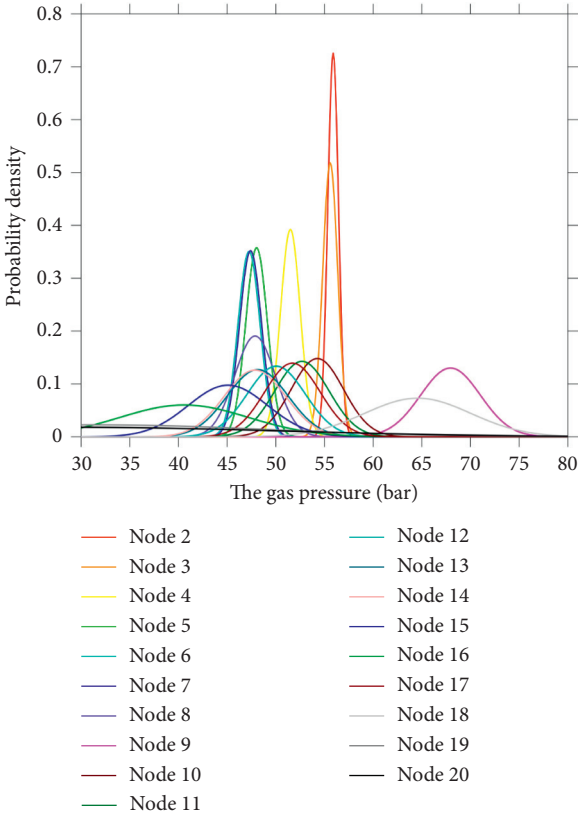


Figure 7: Probability distribution of the eigenvalue.

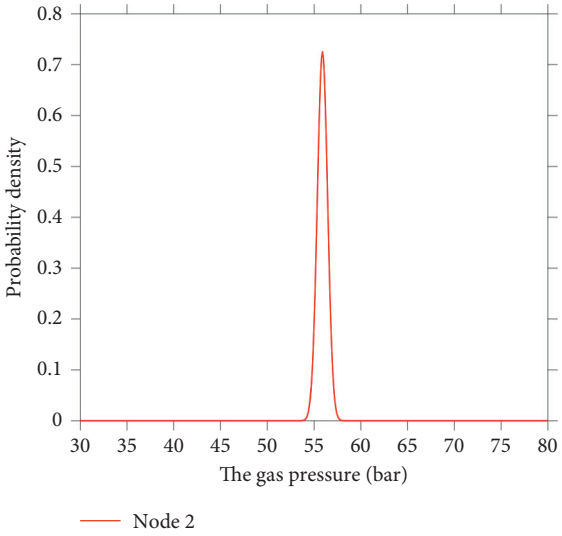


Figure 8: Probability distribution of pressure at node 3.

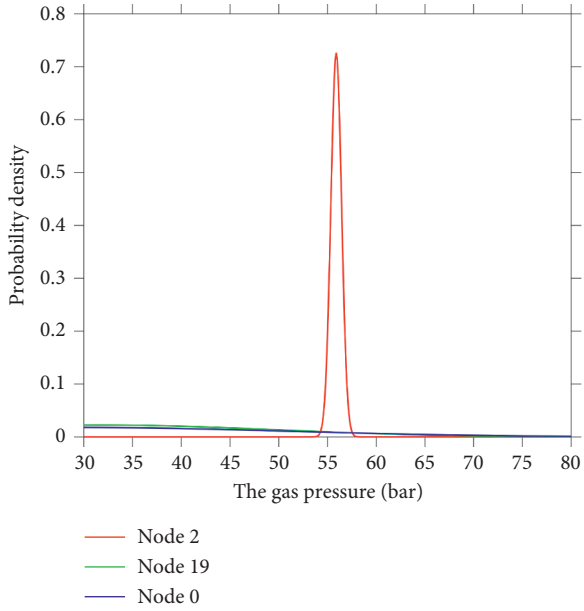


Figure 9: Comparison among Nodes 3, 19, and 20.

stability of the whole gas network, we should pay attention to the stability of the terminal nodes.

6. Conclusion

This paper constructed a coupling network model considering renewable energy generation and P2G technology and did a static stability analysis of this model. This model takes into account the influence of renewable power generation uncertainty on the system. It also takes into account the impact of P2G technology on the gas characteristics specific gravity and gross calorific value of the whole network. In this paper, the static stability of this model is analyzed by probabilistic power flow calculation and characteristic root sensitivity method. The influence of data uncertainty on the stability of power network and gas network is analyzed, respectively. The confidence level of the system can be obtained by determining the confidence interval of power network and gas network. Thus, the potential danger and weak link of the system can be further judged. The feasibility of this method is verified by simulation based on actual data. The effects of the randomness and intermittency of renewable energy on the stability of the system are visually demonstrated.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key R&D Program of China under grant (2018YFA0702200), the Key Projects of National Natural Science Foundation of China (U20A20190), the National Natural Science Foundation of China (62073065), and the Liaoning Province to Revitalize the Liao Talents Plan (XLYC1802046).

References

[1] Q. Sun and L. Yang, "From independence to interconnection-A review of AI technology applied in energy systems," *CSEE Journal of Power and Energy Systems*, vol. 5, no. 1, pp. 21–34, 2019.

[2] Q. Sun, N. Zhang, S. You, and J. Wang, "The dual control with consideration of security operation and economic efficiency for energy hub," *Institute of Electrical and Electronics Engineers Transactions on Smart Grid*, vol. 10, no. 6, pp. 5930–5941, 2019.

[3] Y. Huang, Q. Sun, N. Zhang, and R. Wang, "A multi-slack bus model for bi-directional energy flow analysis of integrated power-gas systems," *CSEE Journal of Power and Energy Systems*, to be published.

[4] N. Zhang, Q. Sun, J. Wang, and L. Yang, "Distributed adaptive dual control via consensus algorithm in the energy internet," *Institute of Electrical and Electronics Engineers Transactions on Industrial Informatics*, vol. 2020, 1 page, 2020.

[5] A. Sawas and H. E. Z. Farag, "Optimal sizing of powerctocgas units toward elevated renewable power penetration," in *Proceedings of the 2019 Institute of Electrical and Electronics Engineers Canadian Conference of Electrical and Computer Engineering (CCECE)*, pp. 1–4, Edmonton, CA, USA, May 2019.

[6] Z. Zeng, T. Ding, Y. Xu, Y. Yang, and Z. Dong, "Reliability evaluation for integrated power-gas systems with power-to-gas and gas storages," *Institute of Electrical and Electronics Engineers Transactions on Power Systems*, vol. 35, no. 1, pp. 571–583, 2020.

[7] T. A. Williams, *Interchangeability of Fuel Gases*, New YorkThe Industrial Press, New York, NY, USA, 1965.

[8] G. Guandalini, S. Campanari, and M. C. Romano, "Power-to-gas plants and gas turbines for improved wind energy dispatchability: energy and economic assessment," *Applied Energy Barking Then Oxford*, vol. 147, pp. 117–130, 2015.

[9] J. Liu, H. Zhong, K. Zeng, H. Fan, and Q. Chen, "Optimal scheduling of multiple energy system considering power to gas unit," in *Proceedings of the 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*, pp. 1–6, Beijing, China, November 2017.

[10] R. Wang, Q. Sun, W. Hu, Y. Li, D. Ma, and P. Wang, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *Institute of Electrical and Electronics Engineers Transactions on Power Electronics*, vol. 2020, 1 page, 2020.

[11] R. Wang, Q. Sun, D. Ma, and Z. Liu, "The small-signal stability analysis of the droop-controlled converter in electromagnetic timescale," *Institute of Electrical and Electronics Engineers Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1459–1469, 2019.

[12] X. Guan, W.-H. E. Liu, and A. D. Papalexopoulo, "Application of a fuzzy set method in an optimal power flow," *Electric Power Systems Research*, vol. 34, no. 1, pp. 11–18, 1995.

[13] G. K. V. Raju and P. R. Bijwe, "Fuzzy distribution power flow for weakly meshed systems," in *Proceedings of the 2007 IEEE Power Engineering Society General Meeting*, p. 1, Tampa, FL, USA, June 2007.

[14] J. M. Morales, L. Baringo, A. J. Conejo, and R. Mínguez, "Probabilistic power flow with correlated wind sources flow with correlated wind sources," *IET Generation, Transmission & Distribution*, vol. 4, no. 5, pp. 641–651, 2010.

[15] A. P. S. Meliopoulos, G. J. Cokkinides, and X. Y. Chao, "A new probabilistic power flow analysis method flow analysis method," *Institute of Electrical and Electronics Engineers Transactions on Power Systems*, vol. 5, no. 1, pp. 182–190, 1990.

[16] K. Qian, C. Zhou, M. Allan, and Y. Yuan, "Modeling of load demand due to ev battery charging in distribution systems," *Institute of Electrical and Electronics Engineers Transactions on Power Systems*, vol. 26, no. 2, pp. 802–810, 2011.

[17] Z. Ren, W. Yan, C. Ding, J. Yu, and X. Zhao, "Probabilistic optimal power flow analysis of virtual power plant containing photovoltaic generation," in *Proceedings of the 2014 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pp. 1–5, Kowloon, Hong Kong, December 2014.

[18] J. Yu, H. Qi, S. Qiu, X. Wang, and H. Zhang, "Probabilistic load flow calculation with irregular distribution variables considering power grid receivability of wind power generation," in *Proceedings of the 2016 IEEE 8th International Power Electronics and Motion Control Conference (IPEMC-ECCE Asia)*, pp. 1457–1461, Hefei, China, May 2016.

[19] Y. Huang and Q. Sun, "Static voltage stability analysis for integrated energy system," *Chinese Journal of Electrical Engineering*, vol. 39, no. 1, pp. 44–53, 2019.

[20] R. Wang, Q. Sun, W. Hu, Y. Li, D. Ma, and P. Wang, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *Institute of Electrical and Electronics Engineers Transactions on Power Electronics*, vol. 2021, 1 page, 2021.

[21] W. Hu, C. Ruan, H. Nian, and D. Sun, "Zero-sequence current suppression strategy with common mode voltage control for open-end winding PMSM drives with common DC bus," *Institute of Electrical and Electronics Engineers Transactions on Industrial Electronics*, vol. 2020, 1 page, 2020.

[22] L. Yang, X. Zhao, X. Hu, G. Sun, and W. Yan, "Probabilistic power and gas flow analysis for electricity-gas coupled networks considering uncertainties in pipeline parameters," in *Proceedings of the 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*, pp. 1–6, Beijing, China, November 2017.

[23] J. Fang, "Dynamic optimal energy flow in the integrated natural gas and electrical power systems," in *Proceedings of the 2018 IEEE Power Energy Society General Meeting (PESGM)*, p. 1, Atlanta, GA, USA, August 2018.

[24] W. E. Dickinson, *Solar Energy Technology Handbook*, Butterworth, London, UK, 1980.

[25] T. Burton, N. Jenkins, D. Sharpe, and E. Bossanyi, *Wind Energy Handbook*, John Wiley and Sons Ltd, Chichester UK, 2001.

[26] A. Shabanpour-Haghighi and A. R. Seifi, "An integrated steady-state operation assessment of electrical, natural gas, and district heating networks," *Institute of Electrical and*

Electronics Engineers Transactions on Power Systems, vol. 31, no. 5, pp. 3636–3647, 2016.

[27] C. M. Correa-Posada and P. Snchez-Martn, "Integrated power and natural gas model for energy adequacy in short-term operation," *Institute of Electrical and Electronics Engineers Transactions on Power Systems*, vol. 30, no. 6, pp. 3347–3355, 2014.

[28] D. De Wolf and Y. Smeers, "The gas transmission problem solved by an extension of the simplex algorithm," *Management Science*, vol. 46, no. 11, pp. 1454–1465, 2000.

Hindawi
Complexity
Volume 2021, Article ID 6645778, 11 pages
https://doi.org/10.1155/2021/6645778

Research Article

An Adaptive Parallel Method for Indexing Transportation Moving Objects

Kun-lun Chen,[1] Chuan-wen Li,[1] Guang Lu,[1] Jia-quan Li,[1] and Tong Zhang[2]

[1]*School of Computer Science & Engineering, Northeastern University, Shenyang 110819, China*
[2]*State Grid Dalian Electric Power Supply Company, Dalian, China*

Correspondence should be addressed to Chuan-wen Li; lichuanwen@mail.neu.edu.cn

Received 28 December 2020; Revised 26 January 2021; Accepted 3 February 2021; Published 22 February 2021

Academic Editor: Rui Wang

Transportation cyber-physical systems are constrained by spatiality and real-time because of their high level of heterogeneity. Therefore, applications like traffic control generally manage moving objects in a single-machine multithreaded manner, whereas suffering from frequent locking operations. To address this problem and improve the throughput of moving object databases, we propose a GPU-accelerated indexing method, based on a grid data structure, combined with quad-trees. We count object movements and decide whether a particular node should be split or be merged on the GPU. In this case, bottlenecked nodes can be translated to quad-tree without interfering with the CPU. Hence, waiting time of other threads caused by locking operations raised by object data updating can be reduced. The method is simple while more adaptive to scenarios where the distribution of moving objects is skewed. It also avoids shortcomings of existing methods with performance bottleneck on the hot area or spending plenty of calculation resources on structure balancing. Experiments suggest that our method shows higher throughput and lower response time than the existing indexing methods. The advantage is even more significant under the skewed distribution of moving objects.

1. Introduction

In modern economic infrastructure, a very important part is the transportation network. It connects cities, manufacturers, retailers, and nations by moving large volumes of freight and passengers through a complex network [1] (highway, railways, and so on). For better utilizing this network, the transportation system generally consists of the physical and cyber system (i.e., transportation cyber-physical systems, TCPSs) to provide innovative services. However, TCPS with mobile devices presents additional complexity because of the uncertainty of the object (vehicles and mobile devices). Many important applications such as traffic control and autonomous driving relay on the moving object satabase (MOD) [2], whose main goal is to provide users with a certain range of query results of mobile objects under the premise of meeting certain time accuracy and spatial accuracy. For example, in taxi applications, mobile users and taxis are mobile objects, and recommending a taxi within a certain distance for users is a typical mobile object query operation [3].

TCPS, in general, demonstrates a high level of heterogeneity, including sensor nodes, mobile devices, high-end workstations, and servers. The different components of TCPS probably have a nonuniform granularity of time and spatiality, TCPS is constrained by spatiality and real-time [4], and traditional single-threaded spatial data management methods are difficult to meet practical application requirements in terms of efficiency for massive mobile object data. Although considerable data processing platforms such as Hadoop and Spark have the advantage in dealing with massive data, their platform structures are mostly distributed and thus required for mutual communication between nodes. The time cost of these platforms cannot meet the real-time requirements of mobile spatial data management. Therefore, mobile object management is currently implemented in a single-machine multithreaded manner.

The facets of TCPS mentioned above require MOD to process large amounts of updates and queries in real-time, and thus spatial indexes are generally used to enhance its performance. There are currently two major categories of mobile object indexing structures: tree-based index and grid-based index. In the beginning, most of the indexes resided in disks since the index itself can be too large to fit in memory. Thus, tree-based indexes [5–7] are more popular structures because of sophisticated improvements that reduce disk I/O. Recently, the rapid development of computer hardware has made low-cost and high-capacity main memory capable of processing millions of moving objects. In fact, due to the high frequency of object position changes, the in-memory data structure for storing moving objects is much better than the on-disk one to improve IO efficiency [8]. Therefore, many works have been done to extend tree-based indexes to in-memory variants, such as [9–11]. However, the tree-based index update operation is specially constructed to reduce disk I/O, thus complicated and time-consuming.

On the other hand, compared with the tree-based index covering the space with its leaf nodes, the grid-based index divides space into uniform grids of uniform size. In the in-memory environment, these simple uniform grids are easy to update and maintain and thus more efficient than its tree-based competitor. Šidlauskas et al. [9] proposed u-Grid, an update efficient grid-based structure where a secondary index is employed to support the bottom-up updates. Xu et al. [12] proposed D-grid that takes advantage of the velocity information for further improving query performance. They also proposed the lazy deletion and garbage cleaning mechanism for accelerated update processing. Šidlauskas et al. [13] proposed PGrid, a parallel main memory indexing technique that supports heavy location-related query and updates operation by exploiting the parallelism of modern processors.

The grid-based index structure indeed is simple and easy to implement, but it is not suitable for uneven distribution, which is widespread in real-world applications [14]. For example, in the traffic monitoring applications, downtown compared with suburban (or in morning and evening peaks compared with other periods), the load on mobile objects' space and time is uneven. Although the tree-based index can make every leaf node contain approximately the same number of moving objects, the experimental results show that its performance is not as good as grid-based indexing [9].

In fact, the performance bottleneck of single-machine multithreaded mobile object management in main memory is no longer the I/O operation but the delay caused by the coordination among multiple threads, the most crucial influence of which is thread's locking operation on data structure on other threads. Reducing locking operations on other threads is the research direction of mobile object management. We observe that the objects need to be locked only when entering and leaving the leaf nodes. Thus, the division of leaf nodes in the tree index should not be determined by the total number of moving objects but the number of objects entering and leaving the leaf nodes in a unit time. However, counting the incoming and outgoing objects involves atomic operations, which will frequently lock the counters. We also need to decide whether to split the leaf nodes or merge the adjacent leaf nodes according to the changing count results that require continuous calculations, which will seriously affect the efficiency of object data update. Therefore, the existing research does not adopt the index method of dividing the leaf nodes by the entry and exit object counts [15].

Therefore, this paper proposes an adaptive parallel GPU-accelerated transportation moving object indexing method, which uses a grid-integrated quad-tree structure and in which GPU process counts the number of incoming and outgoing objects of leaf nodes and whether the leaf nodes need to be split or merged. This method only occupies a small amount of CPU computing resources and achieves the goal of continuously optimizing the index structure without affecting data update and query efficiency. The experimental results show that the performance of this index method is better than that of the best existing method for moving object updating.

2. Index Structure

2.1. Problem Definition. Given space plane S, moving object set $O = \{o_1, \ldots, o_n\}$, any of them $o_i = \{o_i^{id}, o_i^x, o_i^y, o_i^t\}$, o_i^{id} is the unique identifier of o_i, (o_i^x, o_i^y) is its location, and o_i^t is the last update time. Query operation set $Q = \{q_1, \ldots, q_e\}$, in which any query request $q_j = \{x_{\min}, y_{\min}, x_{\max}, y_{\max}, t_q\}$. The first four items define a rectangular query box and t_q is the time when the query started. The purpose of moving the object database is to return the mobile object located in the query box to the user when the query q_j arrives in Q.

This query method is called Range Query. Since other types of queries such as kNN queries can be converted into a series of range queries, this article only needs to discuss the support of the index structure for such queries.

2.1.1. Index Structure Design. The main task of moving the object database is to constantly update the position of the moving object and return the result according to the query requirement. Therefore, the mobile object index structure needs to meet the two basic conditions:

(1) Find object by object identification o_i^{id}

(2) Find and update moving objects based on object position (o_i^x, o_i^y)

An auxiliary index based on the hash table can be used to support condition (1).

Definition 1. Auxiliary index: all spatial objectsusing hash table $\mathscr{H}$ according to o_i^{id} d save key-value pairs like (o_i^{id} *p_bkt*, *idx*) in $\mathscr{H}$. *p_bkt* is the location of the memory space of the bucket in which o_i is located in the hybrid index (see Definition 2). *idx* represents its relative position in the bucket.

Figure 1 shows an example of an auxiliary index. The nature of the hash table shows that using an auxiliary index

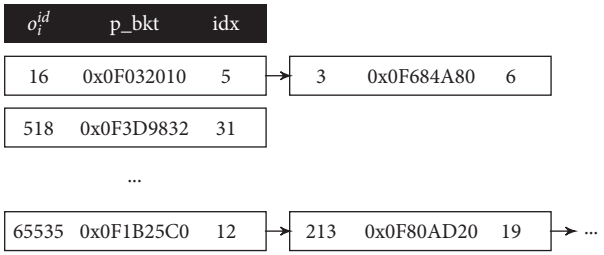


Figure 1: An example of the auxiliary index.

can find the memory location of o_i in memory in a constant time based on o_i^{id}.

Both grid-based and tree-based indexes have advantages and disadvantages in satisfying condition (2). The grid-based indexing method directly calculates the grid to which o_i belongs according to (o_i^x, o_i^y). However, when the moving object space is not uniformly distributed, the number of moving objects is excessive in the grid of hotspot region, and updating the location information of the object will cause the grid in the hotspot region to be locked frequently, which reduces the parallel performance of the hotspot region grid.

Tree-based indexing reduces the congestion of objects in hotspots. However, searching o_i within (o_i^x, o_i^y) requires multiple queries from tree root nodes to a series of intermediate nodes up to leaf nodes, which is inefficient. The more significant effect on overall efficiency is that tree-like indexes need constantly adjusting the structure to fit the distribution of moving objects. Computation of whether leaf nodes needs to be adjusted, and adjust operations themselves will consume a large number of computing resources.

This paper proposes a hybrid indexing method combining grid and quad-tree to avoid the disadvantages of the above two methods.

Definition 2. Hybrid index P divides the space plane S into $G_{\mathrm{num}} = 2^\rho \times 2^\rho$ grids, each of which can be converted between grid nodes and quad-trees depending on the conditions.

Hybrid index balances the load of each cell by transforming the grid of the hotspot region into a quad-tree [16]. ρ is the gridding parameter, using the selection condition given by the document [17]:

$$\rho = \frac{1}{2}\left[\log N - \log C_L\right], \tag{1}$$

where N represents the total number of moving objects in space and C_L represents the capacity of the leaf node.

During execution, the number of moving objects in each grid is dynamically determined, and the grid that satisfies the split requirement is converted into a quad-tree, and the quad-tree that meets the merge requirement is converted back to the grid. Within each quad-tree, its leaf nodes are also split or merged according to conditions.

Figure 2 shows an example of a hybrid index where the uppermost layer in the right half is a plane S that is divided into $G_{\mathrm{num}} = 2^\rho \times 2^\rho$ grids, where the black grid represents the hot grid that is converted to a quad-tree, and the lower

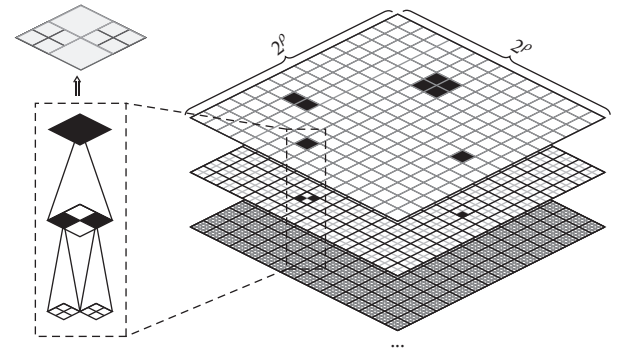


Figure 2: An example of the hybrid index.

layers represent the corresponding spatial regions of the quad-tree nodes at various levels, where the black nodes represent further subdivided regions. The left part of the figure shows one of the zoomed-in quad-trees. The top of the figure shows the quad-tree division to the grid. After converting to a quad-tree, the region is divided more finely.

Each grid node c_i contains only one p_{bucket} pointer to a bucket list $\mathscr{L}_i$. This list contains a series of buckets that hold a fixed number of all moving objects belonging to the node. Each quad-tree node ξ_i is represented as $\xi_i = \{p_{\mathrm{bucket}}, p_{c1}, p_{c2}, p_{c3}, p_{c4}\}$, where p_{c1}, p_{c2}, p_{c3}, and p_{c4} are pointers to a child node in the four quadrants with the current node as a parent.

2.1.2. Node Split Conditions. Since the splitting and merging operations on the grid are essentially the same as those on the quad-tree leaf nodes, we use the term nodes for both conditions.

When an object moves inside a node c, we only need to lock the object itself for updating its position without locking the node. When the moving object enters and exits the node c, the information of this object must be appended or deleted in a bucket corresponding to c and c must be locked simultaneously to prevent data collisions caused by different threads concurrently updating the c. Therefore, the update operation of all other threads to node c needs to be suspended. The greater the number n_c of moving objects into and out of c per unit of time, the more threads waiting at the same time. The total time required for the update operation φ_c is the sum of the actual execution time of the operation $n_c\tau$ (where τ is the execution time of each update operation) and the waiting time ψ_c. Therefore, there is a positive correlation between total update operation time and the number of moving objects in and out of the nodes.

By splitting the nodes, we can reduce the number of simultaneous waiting threads within the node. However, after splitting a node c into nodes, the movement of objects within a node c may become a movement among multiply nodes, increasing the total number of moving objects in and out of the nodes, i.e.,

$$n_c \le n_{c1} + n_{c2} + n_{c3} + n_{c4}. \tag{2}$$

Therefore, to reduce φ, it requires a moderate split of the node, only when

$$\varphi_c > \varphi_{c1} + \varphi_{c2} + \varphi_{c3} + \varphi_{c4}. \tag{3}$$

Splitting the nodes helps to improve update performance. Conversely, when the four split nodes c_1, c_2, c_3, and c_4 satisfy the condition $\varphi_{c1} + \varphi_{c2} + \varphi_{c3} + \varphi_{c4} > \varphi_c$, they need to be merged into one node.

This section will quantify the relationship between φ and n. We will first discuss the effect of each additional update on the total waiting time for moving objects.

Lemma 1. *Set the execution time of the update operation required for each node to enter and exit as τ and the number of moving objects in or out of the node per unit time as n_c ($n_c > 1$). Mark the expected value of the total waiting time required for updating n_c moving objects entry and exit information as $\psi(n_c)$, and then update the expected value $\psi(n_c + 1)$ of the waiting time for $n_c + 1$ moving objects as follows:*

$$\psi(n_c + 1) = \psi(n_c) + \frac{3}{2} n_c \tau^2 + 4(n_c - 1)\tau^3. \tag{4}$$

Proof. Assume that node c has n_c objects into and out of the node within a unit time. Then, the $n_c + 1$ th object cannot move in or out of its cell until the previous $n_c + 1^{\text{th}}$ objects finish updating, which is denoted by the following:

$$\psi(n_c + 1) = \psi(n_c) + \phi(n_c).(n_c \geq 1), \tag{5}$$

where $\phi(n_c)$ is the correlation function between waiting time and the number of moving objects. There are two types of update waiting time among the n_c objects. The overlapping update waiting time means that an object's update start time point appears within the period when another object is performing an update operation, whereas the nonoverlapping update waiting time means that the update start time point of an object occurs before the update start point of another object. The interval between two update start time points is greater than τ. This ensures that two objects are updated without waiting for each other. The significance of the correlation function $\phi(n_c)$ is the relationship between overlapping and nonoverlapping of object updates.

When the update waiting time of n_c objects is not overlapping, we mark the probability that the $n + 1$ object falls within the object execution period as p_1', and $p_1^{"}$ is the probability that $n + 1$ object falls before the time τ before an object updates time in time.

If $p_1' = n_c$, the waiting time at this time is the integral of τ on $[0, \tau]$, i.e.,

$$w_1' = \int_0^\tau \tau \mathrm{d}\tau. \tag{6}$$

If $p_1'' = n_c \times \tau$, the waiting time at this time is the integral of the constant 1 on $[0, \tau]$, i.e.,

$$w_1'' = \int_0^\tau 1 \, d\tau. \tag{7}$$

Finally, the waiting time between n_c objects can be expressed as $w_1 = p_1' \times w_1' + p_1'' \times w_1''$.

When the update waiting time is overlapping among n_c objects, the overlapping time will need to be postponed. The waiting time for the overlapping part is set as w_2, and the execution time of the object is τ. Each object can be overlapped for a period of time τ, so the period that can be overlapped is $o_t = 2\tau$.

The postponed period is denoted as w_2', the integral of the overlapped period o_t on $[0, \tau]$:

$$w_2' = \int_0^\tau 2\tau \, d\tau. \tag{8}$$

There are only overlapping periods among n_c objects, and the postponed period among n_c overlapping objects is

$$w_2' = (n_c - 1) \int_0^\tau 2\tau \mathrm{d}\tau. \tag{9}$$

The concept of overlap between two objects shows that each time the object can be overlapped is denoted as $o_t = 2\tau$. The overlapping probability between two objects is

$$p_2' = 2 \times o_t = 4\tau. \tag{10}$$

The overlapping waiting time w_2 between n_c objects can be expressed as

$$\begin{aligned} w_2 &= p_2' \times w_2', \\ \phi(n_c) &= w_1 + w_2. \end{aligned} \tag{11}$$

Substituting the above formula into it, we get

$$\begin{aligned} \phi(n_c) &= p_1' \times w_1' + p_1'' \times w_1'' + p_2' \times w_2' \\ &= n_c \int_0^\tau \tau \mathrm{d}\tau + n_c \tau \int_0^\tau 1 \mathrm{d}\tau + 4\tau(n_c - 1) \int_0^\tau 2\tau \mathrm{d}\tau \\ &= \frac{3}{2} n_c \tau^2 + 4(n_c - 1)\tau^3. \end{aligned} \tag{12}$$

Finally, $\psi(n_c + 1) = \psi(n_c) + (3/2)n_c\tau^2 + 4(n_c - 1)\tau^3$.

The closed form of the waiting time $\psi(n)$ can be derived from Lemma 1. □

Lemma 2. *Set the number of moving objects into and out of a node per unit time as n and the update execution time of each object as τ. Then, the total waiting time $\psi(n)$ for n update operations is as follows:*

$$\psi(n) = \frac{3}{4} n(n - 1)\tau^2 + 2(\mathrm{n} - 1)(\mathrm{n} - 2)\tau^3. \tag{13}$$

Proof. According to Lemma 1, the update waiting time of a moving object is related to the number of moving objects and the update time point of the object. Formula (4) can be recursively expanded as

$$\begin{aligned}\psi(n) &= \frac{3}{2}(n-1)\tau^2 + 4(\mathrm{n}-2)\tau^3 + \psi(n-1),\\ \psi(n-1) &= \frac{3}{2}(n-1)\tau^2 + 4(\mathrm{n}-3)\tau^3 + \psi(n-2),\\ \psi(2) &= \frac{3}{2}\tau^2 + \psi(1),\\ \psi(1) &= 0.\end{aligned} \tag{14}$$

Substituting $\psi(1), \ldots, \psi(\mathrm{n})$ into the formula, we get

$$\psi(\mathrm{n}) = \sum_{n=2}^{n-1}\frac{3}{2}n\tau^2 + \sum_{n=2}^{n}4(n-2)\times\tau^3. \tag{15}$$

So, Lemma 2 was proved.

Lemma 2 describes the relationship between the object update waiting time and the number of object updates, and the total update time φ is the sum of the update waiting time ψ and the actual execution time $n\tau$. □

Theorem 1. *Set the number of moving objects to and from the node per unit time as n. The update execution time of each object is τ. Then, the total update time $\varphi(n)$ including node waiting for other thread per unit of time can be expressed as follows:*

$$\varphi(n) = n\tau + \frac{3}{4}(n-1)\tau^2 + 2(n-1)(n-2)\tau^3. \tag{16}$$

Proof. Substitute (5) into $\varphi(n) = n\tau + \psi(n)$. Then, Theorem 1 can be proved. □

3. Data Structures and Algorithms

This section uses the C++-like pseudocode to describe our index structure. The index structure is divided into two parts: the auxiliary index and the hybrid index, in which the auxiliary index $\mathscr{H}$ adopts a hash table; i.e., the form is

$$\mathrm{unordered}_{\mathrm{map}}\langle \mathrm{int}, \mathrm{pair}\langle \mathrm{Bucket}*, \mathrm{int}\rangle\rangle. \tag{17}$$

We use this structure to store (o_i^{id}, p_bkt, idx) key-value pairs. The hybrid index takes the form of a 2-D array combined with a quad-tree. Next, we will focus on the hybrid index data structure and its associated algorithms.

The hybrid index data structure includes grid index and QuadTree index, which balances the update load of each cell according to the conditions given by Theorem 1. It merges cells containing a small number of objects to reduce the cost of querying quad-trees. It also split cells that contain larger numbers of objects to reduce update waiting time.

From Section 2.1.2, we know that continuously comparing the relationship between the total execution time φ_c of the node c and the total execution time of its four child nodes $\varphi_{c1} + \varphi_{c2} + \varphi_{c3} + \varphi_{c4}$ is crucial to deciding whether to split or merge a node. Since these comparison operations consume massive computing resources, the indexing method performs these operations on the GPU. Although counting moving objects in or out of nodes does not require much calculation, each time increased, the counter needs to be locked, which seriously impacts the overall parallel performance. Therefore, the counting operation needs to be performed on the GPU.

This chapter introduces the indexing algorithm from two aspects: CPU and GPU. All algorithms are based on a unified data structure stored in the main memory. The data required for GPU operation are copied from main memory to memory only when used. After the calculation result is obtained, the data are copied back to the main memory.

4. Data Structure

The grid index in a hybrid index is implemented using a two-dimensional array:

$$\mathrm{array}\langle \mathrm{array}\langle \mathrm{unique}_{\mathrm{ptr}}\langle \mathrm{Node}\rangle, \mathrm{width}\rangle, \mathrm{height}\rangle. \tag{18}$$

where width and height represent the number of grid columns and rows, respectively. The Node class is the parent of Cell and quad QuadTree. Cell represents a grid that contains only one pointer of type $\mathrm{unique}_{\mathrm{ptr}}\langle \mathrm{Bucket}\rangle$, and QuadTree represents a quad-tree:

```
Struct QuadTree{
unique_ptr⟨Bucket⟩p_bucket;                    / * Bucket Chain Header * /
  array⟨unique_ptr⟨QuadTree⟩, 4⟩children;      / * Child nodes * /
  QuadTree * parent;                           / * Parent nodes * /
  int left, right, floor, ceiling,             / * Coverage * /
}.
```
(19)

Cell or Objects in QuadTree are stored in a linked list bucket. Buckets are used to store moving objects on leaf nodes. The size of the buckets is fixed. The number of moving objects determines the number of buckets of each leaf node. If the bucket is full when the object is inserted, a new bucket is created; if the bucket becomes empty when the object is deleted, the bucket is deleted. The bucket structure is as follows:

```
struct Bucket{
   Site[Max × _size]sites;          / * Array of objects * /
   int current,                     / * The number of objects in the bucket * /
   unique_ptr⟨Bucket⟩bucket,        / * Next bucket * /
}.
```
(20)

The Site class holds moving object data, including id, x, y and updates time t_u. It can be inferred from the contents of the Site class that the data of one object needs at least 128 bits of memory space (four int values on a 32 bit machine). An access violation may occur between different threads when reading and writing Site classes in parallel without protection. The traditional way to avoid an access violation is to lock it while reading and writing Site. Since Site is the most frequently used class in the index, in order to avoid the impact of locking on performance, we merge the four data in the Site class into an object of type _m128i, use _mm_load_si128, and _mm_set_epi32 operation in the Intel MMX instruction set [18] to read and write the content and use _mm_extract_epi32 to extract the corresponding data. In this way, the index can correctly read and write Site data without locking.

The index also maintains two ⟨Node *⟩ lists: split_candicates list and merge_candidates list. Both save nodes need splitting and need merging separately. We calculate the nodes that need splitting or merging on the GPU.

4.1. CPU Algorithm. As the object continues to update, the structure of the quad-tree will change. The object's insertion algorithm inserts it into the appropriate node based on its coordinates. The deletion algorithm finds the bucket based on the object id and deletes it. The division or merging of cells is a crucial operation for balancing quad-trees. Only cells that meet the conditions for splitting and merging can be divided and merged.

4.1.1. Spatial Object Insertion Algorithm. During the movement of the object, as the position changes, the object will continue to move between nodes. The main purpose of Algorithm 1 is to insert an object into the leaf node. The idea is to find the leaf node to which the object should be inserted based on the position of the object (o_i^x, o_i^y) and determine the bucket state of the node. If the bucket is not full, insert the bucket directly. If the bucket is full, create a new bucket n_bucket and insert the new bucket into the bucket list of the current node. Then, insert the object o_i into the bucket and increase the number of objects stored in the bucket by one.

4.1.2. Spatial Object Deletion Algorithm. When the location of a moving object is updated, objects that no longer belong to the current leaf node range need inserting into its new belonging leaf node and deleted from the current node.

As shown in Algorithm 2, it finds the bucket where the object is located according to the object that uniquely identifies (o_i^{id}). The object is then removed from the bucket, and the number of objects stored in the bucket is decremented. After deleting the object o_i, if the bucket is empty, the bucket will also be deleted.

4.1.3. Cell Partitioning Algorithm. According to the node splitting condition in Section 2.1.2, when the leaf node satisfies the condition $\varphi_c > \varphi_{c1} + \varphi_{c2} + \varphi_{c3} + \varphi_{c4}$, the Quad-Tree index structure should divide the leaf node s_node into four equal-sized subgrids, where x_middle = (left + right/2) and y_middle = (floor + ceiling/2) (lines 1-2). After the division is completed, all children's parent nodes are set as the current leaf nodes, and all the objects in the divided node s_node are moved to the bucket of the corresponding child. If the parent of the partitioned node belongs to the split_candidates list before the node is unpartitioned, it will be removed from the list split_candidates and appended the current node to split_candidates list as shown in Algorithm 3.

4.1.4. Cell Merge Algorithm. According to the node split merge condition in Section 2.1.2, when the leaf node satisfies the condition $\varphi_{c1} + \varphi_{c2} + \varphi_{c3} + \varphi_{c4} > \varphi_c$, QuadTree index structure merges four grids into one. The purpose is to adjust the quad-tree structure and balance the update load to improve the efficiency of the update query.

As shown in Algorithm 4, when a node satisfies the merge condition, buckets of all children of the node are linked and assigned to the bucket of the current node. Adjust the bucket chain of the current node and delete the empty bucket. If the current node belongs to the merge_candidates list, the current node m_node is deleted from merge_candidates. After merging, if the parent node of the current node belongs to the merge_candidates list, the parent node m_node.parent of the current node is added to merge_candidates.

4.2. GPU Algorithm. GPU's logic computation ability is weaker than that of the CPU, so it is not suitable for making complex logical decisions and is more suitable for a large amount of parallel data calculation. Different from CPU, GPU adopts a SIMD (single-instruction, multiple-data) operating mode. The same instruction can be executed on multiple sets of data at the same time so as to improve the efficiency of parallel execution [19]. Therefore, the work that is handed to the GPU in the index is mainly composed of two parts that involve large data amount calculations: (i) counting the number of moving objects entering and leaving the node and (ii) deciding whether the nodes need to be split or merged. Both operations involve

```
Input: move the object o_i = {o_i^id, o_i^x, o_i^y, o_i^t}
Output: no output; the leaf node is updated after the operation is completed
(1)  cur_leaf = get_leaf (o_i^x, o_i^y) /* find the inserted leaf node by location (o_i^x, o_i^y) */
(2)  if (is_fill(cur_leaf.p_bucket))
(3)     n_bucket = new Bucket();
(4)     insert_bucket (n_bucket);
(5)  Insert_object(o_i); /* insert the object o_i into the bucket of the current node */
```

Algorithm 1: Object o insert leaf node add_to_leaf.

```
Input: move the object o_i = {o_i^id, o_i^x, o_i^y, o_i^t}
Output: no output; the leaf node is updated after the operation is completed
(1)  bucket = get_bucket (o_i^id);
(2)  delete_from_bucket (o_i);
(3)  if (is_empty(bucket))
(4)     delete_bucket(bucket); /* delete empty bucket */
```

Algorithm 2: Deleting object from leaf node O remove_from_leaf.

```
Input: quad-tree leaf node
s_node = {p_bucket, children, parent, left, right, floor, ceiling}
Output: no output; quad-tree structure changes after the operation
(1)  x_middle = (left + right)/2;
(2)  y_middle = (floor + ceiling)/2;
(3)  s_node.children[0] = new QuadTree (x_middle, right, y_middle, ceiling; /* in a similar way to initialize s_node.children [1–3] */
(4)  for each (child in s_node.children)
        child.parent = s_node;
           s_node.p_bucket. o_i ⟶ child.p_bucket;
(5) if (s_node.parent in split_candidates)
        delete_from_split_candidates (s_node.parent);
(6)  insert_split_candidates (s_node);
```

Algorithm 3: Cell division split.

```
Input: quad-tree node: m_node = {p_bucket, children, parent, left, right, floor, ceiling}
Output: no output; quad-tree structure changes after the operation
(1)  if(for each m_node, children is_leaf()) /*all children with the node m_node are leaf nodes*/
(2)     m_node.p_bucket ← child.p_bucket;
(3)     delete _null_bucket (m_node.p_bucket);
(4)     if (m_node in merge_candidates)
(5)        delete _from_merge_candidates (m_node)
(6)     if (for each m_node.parent.child is_leaf())
(7)        insert _merg_candidates (m_node.parent);
```

Algorithm 4: Cell merge merge.

two ⟨Node ∗⟩ type lists split_candidates and merge_candidates, which are maintained in the index.

4.2.1. Counting Algorithm. We take counting objects entering the node as an example to describe the counting algorithm in this section. If the operation runs on the CPU, several threads are usually opened according to the parallel capability of the CPU, and the following processes are executed, respectively, Algorithm 5 is as follows.

update_info_list stores a series of object movement data received by the system, including the object id, original location, and new location. Each CPU thread handles part of the update_info_list. After the node where the object is moved is calculated by get_location, and the third row increments the shifted-in node counter. Note that the counter

```
Input: partial object movement information linked list update_info_list
Output: update each node counter value
(1)   for each (item in some part of update_info_list)
(2)     node_in = get_location (item.new_pos);
(3)     lock_and_increase (node_in);
```

Algorithm 5: CPU-based counting algorithm CPU_Count.

must be locked when incrementing to ensure correct reading and writing of counters in parallel.

Compared with the CPU, GPU generally has thousands of stream processors [13], each of which maintains a fixed number of counters for several nodes. Since each counter can only be accessed by a fixed stream processor, it does not need to be locked when incrementing. Algorithm 6 demonstrates that each stream processor is responsible for $m \times n$ node operation flow.

threadIdx is a built-in stream handler index value automatically set by the system. Count array is a local array of stream processors and saves the node counters it is responsible for. Line 6 determines whether the node belongs to the node that the stream processor is responsible for, and if it belongs to the stream processor, it increments the corresponding counter (line 7).

The indexing system not only calls GPU_count on all nodes in the lists split_candidates and merge_candidates to record the number of objects in or out of nodes but also calls GPU_count on four child nodes that each node may divide in the list split_candidates and parent nodes that may be merged by four neighboring nodes in the list merge_candidates.

4.2.2. Split Merge Judgment Algorithm. We take the split judging algorithm as an example. Because the indexing system has counted any node c in the split_candidates list and the four child nodes c_1, c_2, c_3, and c_4 in which it may be split into, we just need to assign nodes in the split_candidates list to each stream handler and then substitute equation (16) into (3) to make the decision.

5. Simulation Experiment and Result Analysis

This section compares the index structure, denoted as GAPI with PGrid. PGrid is the state-of-the-art parallel moving object index structure and is widely adopted in applications for indexing moving objects [13].

The experimental simulation environment is Win 10 system, Intel Xeon E5-2620 v3 six-core CPU × 2, NVIDIA Quadro K2200 GPU. The experimental code is implemented by *C++* and CUDA Toolkit 7.5, and the spatial object data are generated by the open-source mobile object generation tool MOTO (http://moto.sourceforge.net) based on the Brinkhoff [20] algorithm. We set the parameter of GAPI as $\rho = (1/2)[\log N - \log C_L]$ according to equation (1), where N represents the total number of moving objects in space and C_L represents the capacity of the leaf node. Experimental parameters are shown in Table 1.

Figure 3 shows the experimental results of the effect of the update/query ratio on throughput. With the increase in the update/inquiry ratio, the throughput of the two index structures increases. Among them, our method's throughput is higher than that of PGrid in most cases but is only lower when the update/query is small. This is because our method mainly optimizes the update operation. The GPU-assisted splitting and merging of the quad-tree reduce the thread waiting time in the parallel update. Therefore, the average update time is better than PGrid. Because the query on the quad-tree structure needs multiple queries from the root node to the leaf nodes, our index's query time is generally higher than that of the PGrid.

Figure 4 shows the experimental results of the effect of query area on throughput. From the figure, our method is better than PGrid in all cases. With the increase in the area of the query area, the overall throughput is declining. This is because the query operation occupies more and more computing resources.

Figure 5 shows the experimental results of the effect of the update interval on throughput. The longer the update interval, the greater the probability that the object moves out of the current node. When the renewal interval gradually increases and the PGrid throughput decreases drastically, the throughput of our method can still be stable. This is because our method can dynamically adjust the nodes according to the actual situation at running time. We can also see that the dropping speed of GAPI is getting slower as update intervals increase. The reason is that a larger update interval allows GAPI to have enough time to adjust its nodes dynamically, which can balance the effect of the larger moving out ratio and get a rather stable throughput. Figure 5 also shows that our throughput is better than PGrid in all cases.

Figure 6 shows the experimental results of the effect of the number of moving objects on throughput. As the number of moving objects increases, the throughput of both types of indexes decreases. This is because the greater the number of moving objects, the greater the probability of simultaneous updates of multiple threads in the same node and the more threads that need to wait. However, as the number of mobile objects increases, the throughput of our method decreases significantly slower than PGrid. This is because our method can dynamically split the nodes according to the actual running

```
Input: object movement information list update_info_list
Output: update each node counter value
(1)  for each (item in update_info_list)
(2)    node_in = get_location (item.new_pos)
(3)    if (node_in.x/m) == threadIdx.x and
(4)      node_out.y/n = = threadIdx.y
(5)      increase (count[node_in.x%threadIdx.x] [node_in.y%threadIdx.y]);
```

Algorithm 6: GPU-based counting algorithm GPU_count.

Table 1: Experiment parameters.

Parameter	Experimental value	Defaults
The total area (km^2)	200000×200000	—
Query area (km^2)	0.25, 1, 4, 16, 32	4
CPU threads	24	24
GPU stream processors	640	640
Update/query ratio ($\times 10^3$)	0.25, 0.5, 1, 2, 4, 8, 16	1
Number of spatial objects ($\times 10^6$)	5, 10, 20, 40	10
Update interval (second)	10, 20, 40, 80, 160	10

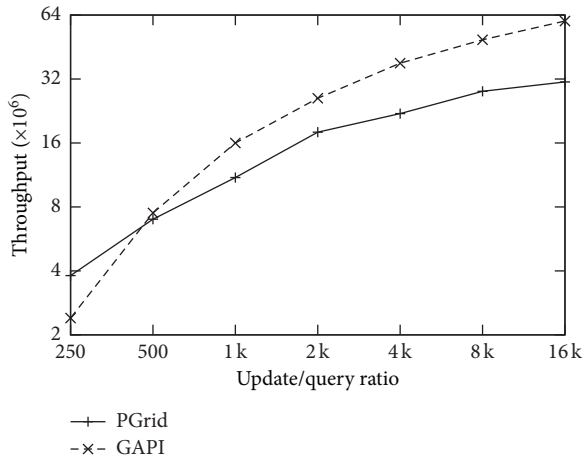


Figure 3: The effect of throughput on update/query ratio.

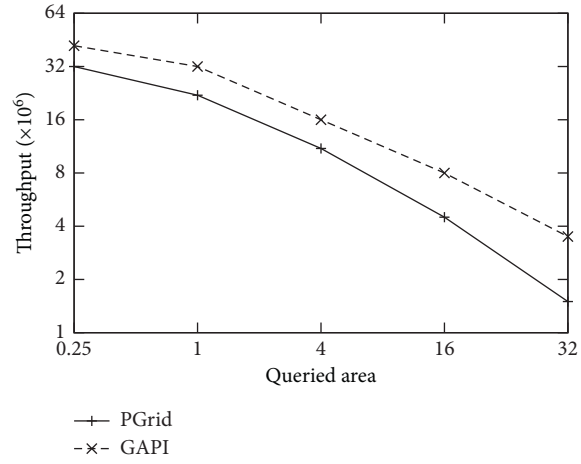


Figure 4: The effect of throughput on the size of the queried area.

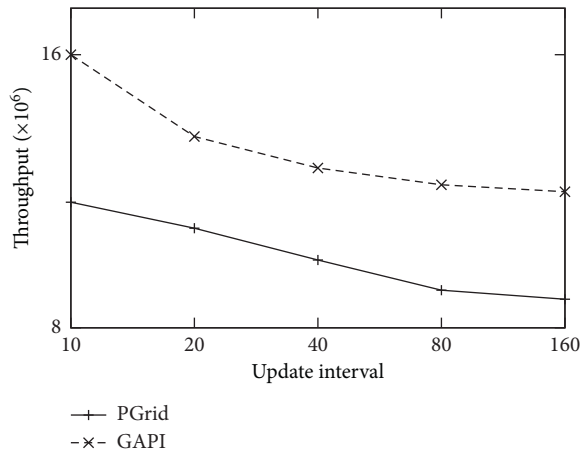


Figure 5: Effect of throughput on the interval between updating.

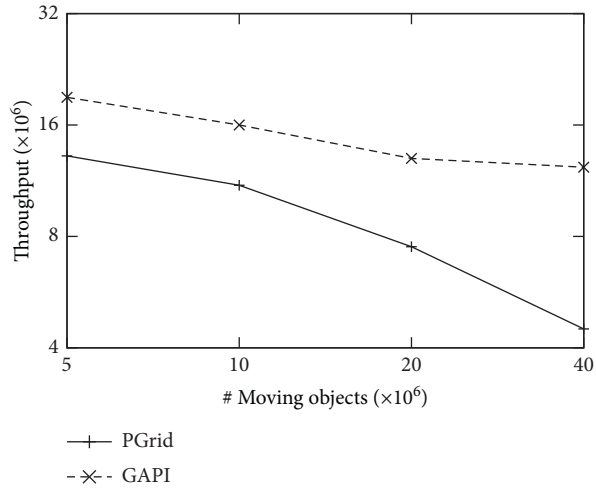


Figure 6: Effect of throughput on the number of moving objects.

conditions, which reduces the number of objects in the nodes, reduces the number of threads that need to wait, and improves the efficiency of parallel updating.

6. Conclusions

For satisfying the requirement of TCPS applications' spatiality and timeliness, based on the grid index, this paper proposes a hybrid indexing method that combines quad-tree and GPU acceleration to avoid the disadvantages of tree-based indexing and grid-based indexing. And, through experimental verification, the following conclusions have been drawn:

(1) Our index passes the calculation of the balanced index structure to the GPU for processing, making full use of the advantage that the GPU can quickly calculate a large amount of data, reducing the CPU's computational load as much as possible, thereby greatly improving the index optimization efficiency.

(2) The index structure divides the number of objects entering and leaving the leaf nodes per unit of time into criteria. Compared with the traditional dynamic index structure that uses the number of leaf nodes as the dividing criteria, the index structure has a better balance of hotspots to update the load. Capability performance in moving object updates is significantly better than existing methods.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they haveno conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Key R&D Program of China (2018YFB1003404) and the National Nature Science Foundation of China (61872071).

References

[1] D. P. F. Möller and H. Vakilzadian, "Cyber-physical systems in smart transportation," in *Proceedings of the 2016 IEEE International Conference on Electro Information Technology (EIT)*, Grand Forks, ND, USA, 2016.

[2] Y.-K. Huang, "Indexing and querying moving objects with uncertain speed and direction in spatiotemporal databases," *Journal of Geographical Systems*, vol. 16, no. 2, pp. 139–160, 2014.

[3] T. Nguyen, Z. He, R. Zhang, and P. Ward, "Boosting moving object indexing through velocity partitioning," in *Proceedings of the VLDB Endowment*, vol. 5, no. 9, pp. 860–871, 2012.

[4] G. Xiong, "Cyber-physical-social system in intelligent transportation," *IEEE/CAA Journal of Automatica Sinica.*vol. 2, no. 3, pp. 320–333, 2015.

[5] S. Saltenis, C. S. Jensen, S. T. Leutenegger, and M. A. Lopez, "Indexing the positions of continuously moving objects," *SIGMOD Conference*, vol. 2, pp. 331–342, 2000.

[6] Y. N. Silva, X. Xiong, and W. G. Aref, "The rum-tree: supporting frequent updates in r-trees using memos," *The VLDB Journal*, vol. 18, no. 3, pp. 719–738, 2009.

[7] Y. Zhu, S. Wang, X. Zhou, and Y. Zhang, "Rum ± tree: a new multidimensional index supporting frequent updates," *Web-Age Information Management*, vol. 2, pp. 235–240, 2013.

[8] D. Idlauskas, S. Altenis, and C. S. Jensen, "Processing of extreme moving-object update and query workloads in main memory," *The VLDB Journal—The International Journal on Very Large Data Bases*, vol. 23, no. 5, pp. 817–841, 2014.

[9] S. S. Sidlauskas, J. M. J. Christiansen, and D. Saulys, "Trees or grids?: indexing moving objects in main memory," *SIGSPATIAL*, vol. 2, pp. 236–245, 2009.

[10] X. Xu, L. Xiong, V. Sunderam, J. Liu, and J. Luo, "Speed partitioning for indexing moving objects," *Advances in Spatial and Temporal Databases*, vol. 2, pp. 216–234, 2015.

[11] R. B. Ray and A. K. Goel, "Supporting location-based services in a main-memory database," *MDM*, vol. 2, pp. 3–12, 2014.

[12] X. Xu, L. Xiong, and V. Sunderam, "D-grid: an in-memory dual space grid index for moving object databases," *17th IEEE International Conference on Mobile Data Management (MDM)*, vol. 1, pp. 252–261, 2016.

[13] D. Šidlauskas, S. Šaltenis, and C. S. Jensen, "Parallel main-memory indexing for moving-object query and update workloads," *SIGMOD*, vol. 2, pp. 37–48, 2012.

[14] Q. Che, C.-W. Li, Y. Zhang et al., "GAPI: GPU accelerated parallel method for indexing moving objects," *Journal of Frontiers of Computer Science and Technology*, vol. 11, no. 11, pp. 1713–1722, 2017.

[15] L.-V. Nguyen-Dinh, W. G. Aref, and M. Mokbel, "Spatio-temporal access methods: Part 2 (2003-2010)," *IEEE Data Engineering Bulletin*, vol. 33, no. 2, pp. 46–55, 2010.

[16] J. Tang, Z. Zhou, K. Ning et al., "A novel spatial indexing mechanism leveraging dynamic quad-tree regional division," 2013.

[17] S. Chen, B. C. Ooi, K.-L. Tan et al., "ST 2 B-tree: a self-tunable spatio-temporal b ± tree index for moving objects," 2008.

[18] A. Peleg, S. Wilkie, and U. Weiser, "Intel MMX for multimedia PCs," *Communications of the ACM*, vol. 40, no. 1, pp. 24–38, 1997.

[19] S. Cook, *CUDA Programming: A Developer's Guide to Parallel Computing with GPUs*, Newnes, London, UK, 2012.

[20] A. D. Sarma, S. Gollapudi, M. Najork et al., "A sketch-based distance oracle for web-scale graphs," 2010.

Hindawi
Complexity
Volume 2021, Article ID 6621235, 22 pages
https://doi.org/10.1155/2021/6621235

Review Article

Industry 4.0-Driven Development of Optimization Algorithms: A Systematic Overview

Róbert Csalódi,[1] **Zoltán Süle**,[1,2] **Szilárd Jaskó**,[1,3] **Tibor Holczinger**,[1,3] **and János Abonyi**[1]

[1]*MTA-PE Lendület Complex Systems Monitoring Research Group, University of Pannonia, Veszprém H-8200, Hungary*
[2]*Department of Computer Science and Systems Technology, University of Pannonia, Veszprém H-8200, Hungary*
[3]*Department of Applied Informatics, University of Pannonia, Nagykanizsa Campus, Nagykanizsa H-8800, Hungary*

Correspondence should be addressed to János Abonyi; janos@abonyilab.com

Received 17 December 2020; Accepted 30 January 2021; Published 13 February 2021

Academic Editor: Renke Han

The Fourth Industrial Revolution means the digital transformation of production systems. Cyber-physical systems allow for the horizontal and vertical integration of these production systems as well as the exploitation of the benefits via optimization tools. This article reviews the impact of Industry 4.0 solutions concerning optimization tasks and optimization algorithms, in addition to the identification of the new R&D directions driven by new application options. The basic organizing principle of this overview of the literature is to explore the requirements of optimization tasks, which are needed to perform horizontal and vertical integration. This systematic review presents content from 900 articles on Industry 4.0 and optimization as well as 388 articles on Industry 4.0 and scheduling. It is our hope that this work can serve as a starting point for researchers and developers in the field.

1. Introduction

The Fourth Industrial Revolution aims to increase the efficiency of production systems and exploit the advantages of digital transformation. Digital transformation is built on cyber-physical systems, which are complex objects that are created by connecting different types of elements (IT, mechanical, etc.) and can interact with physical entities [1]. They facilitate monitoring as well as interactions with other components and can be adapted to production processes [2]. Cyber-physical systems are one of the key enabling technologies of Industry 4.0. Cyber-physical systems are able to effectively support the optimization of production processes, which helps future requirements to be met [3].

The biggest challenges and opportunities caused by the Fourth Industrial Revolution are introduced in Figure 1. It is quite visible that the basis of competitiveness in this new production model is personalised production, which requires the enhanced flexibility and complexity of production processes driven by growing product variety. Obviously, this greater complexity requires solutions to more and more challenging optimization issues. These issues are listed in Table 1.

As the development of Industry 4.0 solutions is related to optimization tasks, this paper aims to overview the optimization tasks, algorithms, and future development direction of these systems according to the Industry 4.0 concept.

General review papers about cyber-physical systems as an enabling technology of Industry 4.0 can be found [4]. Although an enormous number of information sources can be used, no review papers have been written about the optimization solutions of Industry 4.0. The goal, inspired by this shortcoming, is to publish a review paper targeting the topic of optimization in Industry 4.0, which is suitable as a starting point for researchers and developers in the field. The need for this kind of focused analysis has been proven by the publication of many targeted review articles over recent years, e.g., on topics such as the influence of Industry 4.0 on energy consumption [5], the enhancement of fault diagnosis by machine learning methods [6], the combination of opportunities of Industry 4.0 and lean philosophy [7], and the handling of cybersecurity risks [8]. The concept of this work

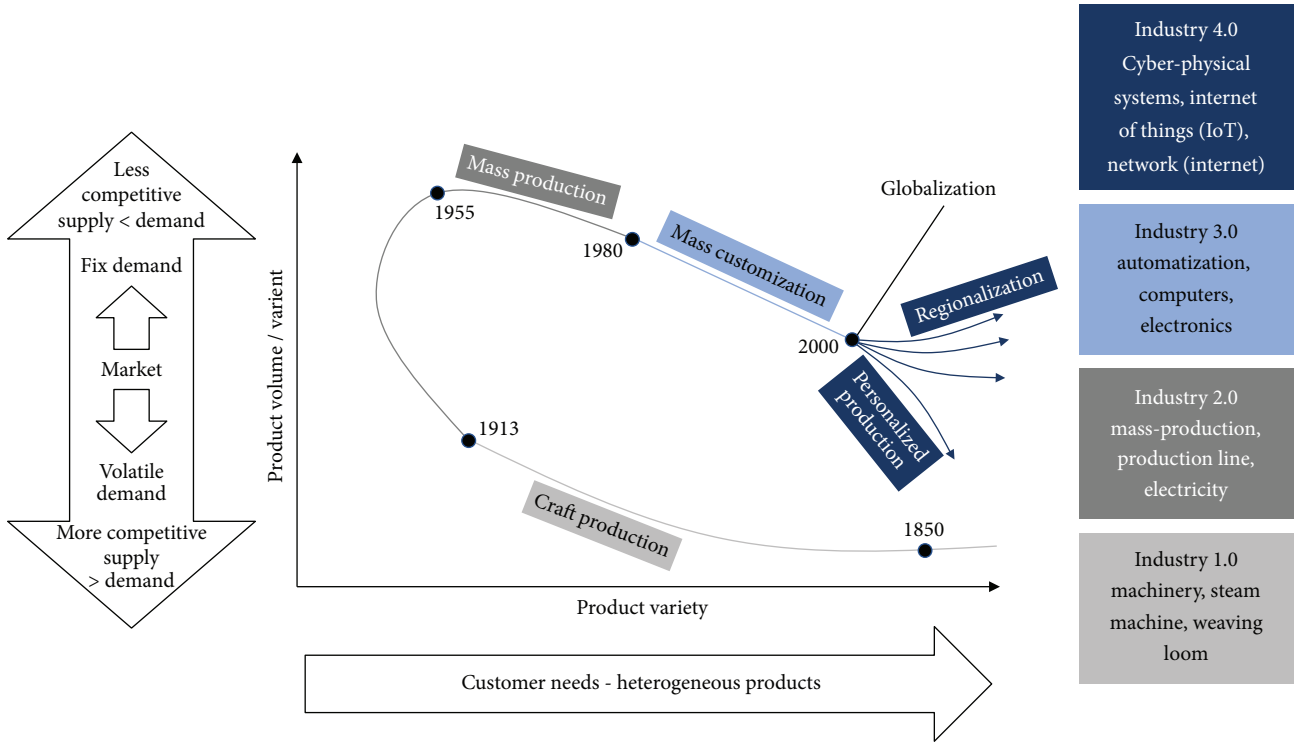


Figure 1: The evolution of industry and the change in drivers.

Table 1: Optimization issues in Industry 4.0 systems.

Objectives	Restrictions	Manufacturing operations
Determine the size of a job (batch, run)	● Inventory	● Highly complex
Assign it to a production line	● Availability of resources (labor, machines)	● Rely on manual scheduling/use complex spreadsheets
Sequence jobs on each line	● Changeovers/cleanings/planned maintenance	● Need to improve on-time delivery
✓Maximize throughput/quantity produced	● Unplanned downtime	● Experience inefficient changeovers/setups
✓Minimize operating cost/ changeovers	● Operator availability	● Have multiple products sharing common resources (labor, infrastructure)
✓Maximize on-time delivery	● Emergency/expedited customer orders	● Production costs are a large proportion of the cost of goods sold
✓Minimize total time to complete production	● Maintenance requirements ● Delivery promises	● Want to increase production without additional capex ● Need to reduce operating costs, product losses/waste

is based on the requirement to develop Industry 4.0 solutions to systematize the publications related to optimization tasks.

This systematic review is based on an examination of the literature available from Scopus, following the PRISMA-P (Preferred Reporting Items for Systematic reviews and Meta-Analysis Protocols) [9]. The PRISMA-P workflow consists of a 17-item checklist intended to facilitate the preparation and reporting of a robust protocol for systematic reviews. The information sources were last fully queried in October 2020 with the following keywords:

(1) (Industry 4.0 AND (optimization OR optimisation))

(2) (Industry 4.0 AND scheduling)

The studied papers were published between 2013 and 2021 and consist of 900 articles on optimization and Industry 4.0 as well as 388 articles on scheduling and Industry 4.0 topics. The complexity of optimization tasks and the approaches to solve them are highly varied. Deviations are caused as a consequence of horizontal and vertical integration. The key contributions and structure of this paper are as follows:

(1) The requirements arising from horizontal and vertical integration that determine the development of optimization algorithms are introduced in Section 2.1.

(2) The characteristics of Industry 4.0 solutions are summarized and the requirements for optimization tasks are presented in Section 2.2.

(3) The quantitative analysis of the 900 + 388 articles related to optimization and scheduling, respectively, is performed based on the aspects explored, the typical applications determined, and the relationship system of the solutions described by network analysis in Section 3.

(4) Typical optimization tasks are formalized, their complexity and metric system are determined, and the impact of changes on the development of these algorithms inducted by the analysis of Industry 4.0-driven solutions is analyzed in Section 4.1.

(5) The characteristics of optimization tasks and the requirements of Industry 4.0 solutions are presented in Section 4.2.

(6) Future research areas are identified based on current trends in Section 4.3.

2. Horizontal and Vertical Integration of Industry 4.0 and the Related Requirements

The purpose of this chapter is to describe the principles of developing Industry 4.0 solutions before presenting optimization tasks. It aims to identify the driving forces that may determine the development of optimization algorithms. Vertical integration and horizontal integration as the main defining factors are discussed in Section 2.1, while features and further requirements of Industry 4.0 are introduced in Section 2.2.

2.1. Horizontal and Vertical Integration of Industry 4.0. The complexity and nature of optimization problems are primarily varied due to horizontal and vertical integration established by Industry 4.0. The purpose of this section is to present these trends and features. Optimization tasks are executed in various fields of Industry 4.0. Improving the efficiency of the Production and Operations Management (POM) is also a cardinal task to increase performance. Two forms of integration approaches are designed, namely, the horizontal and vertical integration. Both types, which are discussed in the following sections, have special requirements and aim to support the POM.

2.1.1. Requirements of Vertical Integration. Vertical integration focuses on the internal integration of manufacturing enterprises and aims to convert data, events, and information from the real world into the digital world, and vice versa [10]. The model is described by the ISA-95 standard [11] and was established to support the development of POM information systems. This standard defines five levels of manufacturing tasks [12], and its goal is to make the connection between the lowest (production execution) and highest (business planning and logistics) stages by integration. The levels and descriptions of ISA-95 can be seen in Figure 2, and various optimization and decision-making exercises are also introduced. The levels define different types of optimization and decision tasks [13]. It can be observed, for example, at the third and fourth levels that operation needs to be optimized by analyzing the adequacy of regulatory and control processes. Schedules and production plans can be qualified by the indicators of appropriateness in order to set suitable input parameters. Planning, scheduling, and optimization are inputs of each other, forming a set of closed loops, and therefore depend on each other. In an integrated solution, it is advisable to solve these optimization tasks simultaneously, in an integrated way.

In summary, the requirements of vertical integration are as follows:

(1) As planning, scheduling, and operation tasks are dependent on each other, these tasks should be solved simultaneously and integrated

(2) More and more information should be extracted to provide informative feedback to support real-time optimization

2.1.2. Requirements of Horizontal Integration. Horizontal integration aims to bring together all the supplier and customer networks, thereby supporting supply chain management. Therefore, information on the production processes is acquired and analyzed in real time in an intelligent production network system by connecting with the Global Enterprise Cloud [14]. This cloud involves the individual entities in the partnership and optimization should cover the entire supply chain within and between organizations simultaneously. This network is called a connected supply chain, and an example of its construction is presented in Figure 3. A connected supply chain will lead to an efficient operation and reduces the costs dramatically but increases complexity.

The most relevant example of horizontal integration is described by collaborative manufacturing management (CMM) [16]. CMM is the practice of organizing and managing manufacturing enterprises. It focuses on the collaboration between partners, Business Process Management (BPM) services, and real-time strategic business management tools. It connects internal business and manufacturing processes as well as synchronizes them with external business processes. It focuses primarily on facilitating in addition to managing business processes and secondarily on support systems [17]. Through collaboration skills, capacities and capabilities are offered to companies by sharing resources and essential information [18]. Actually, CMM is not a new concept. The necessity of this type of integration has already been recognized, but, at the time, could not be properly developed. It was hard to find a suitable partnership to develop a nonhierarchical network, but Industry 4.0 tools seem to support the CMM structure [19]. A network needs to be organized that integrates the various customer-supplier and design-support aspects most effectively in order to increase the efficiency of the business and production [20]. This can be framed by establishing appropriate internal and external links between partners and within the company. The relationship between these requirements is considered in Figure 4. Equipment connectivity aims to join production systems; therefore, all the

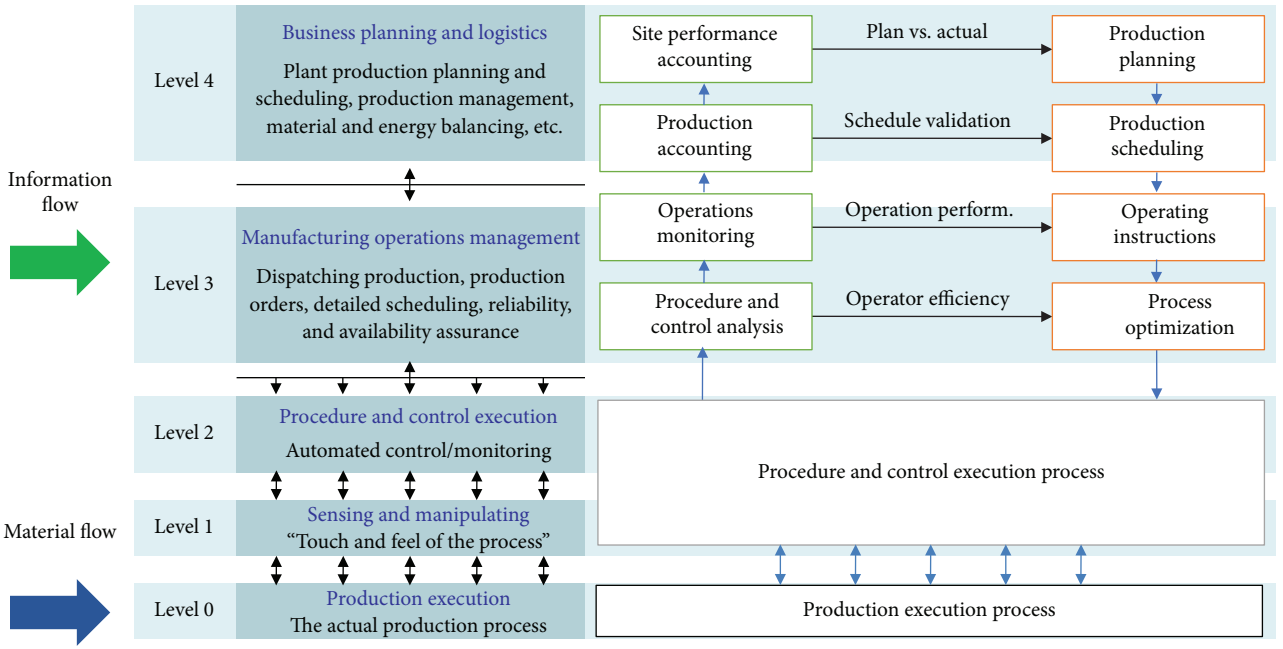


Figure 2: The structure of the ISA-95 standard for constructing vertical integration. It defines 5 levels for the internal hierarchy of manufacturing enterprises. Its purpose is to bridge the highest and lowest stage. A description can be found on the left-hand side of the figure, while optimization and decision-making exercises are also introduced on the right-hand side.

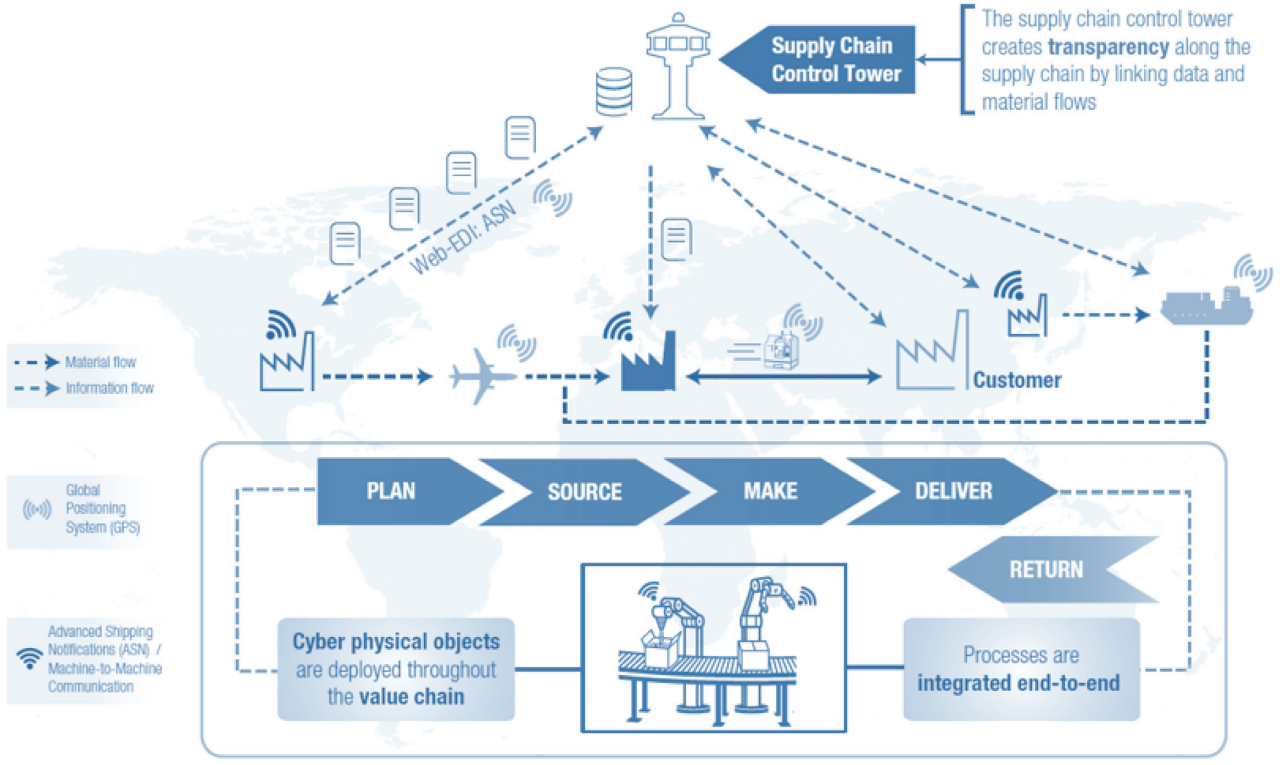


Figure 3: The structure of horizontal integration. Components of the partnership can be found all over the world. An organized supply chain is formed for tracking the life cycle of the product. Planning, sourcing, manufacturing, and delivery processes operate simultaneously. Cyber-physical systems and their connectivity via GPS or Internet enable full communication in order to set the optimal parameters [15].

information is available to the elements of the network. Visibility requires the aggregation of data about the operation to define performance indicators in order to qualify the system appropriately.

The efficient operation of a CMM structure requires complex tasks to be solved. The available studies are about hypernetwork-based model scheduling to improve the utilization and efficiency of manufacturing [21]; the

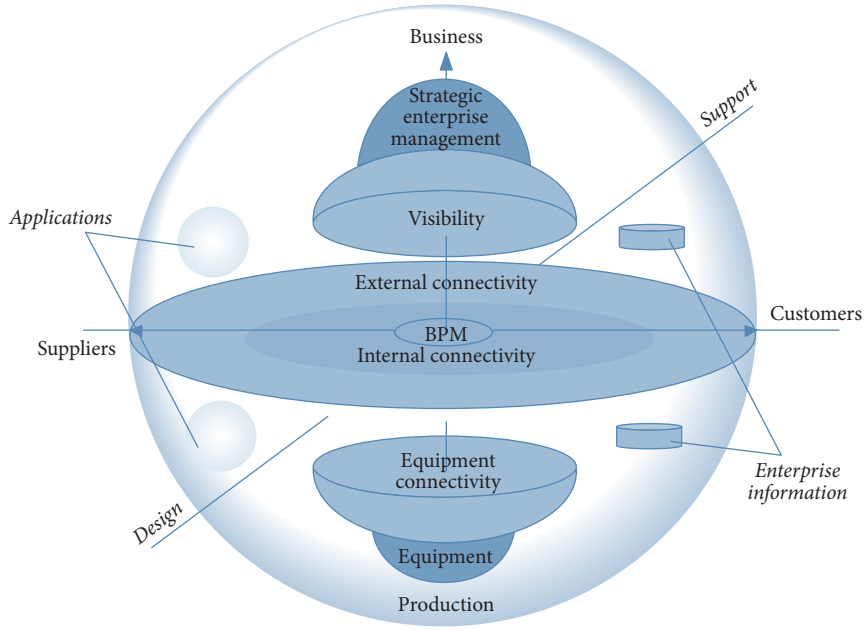


Figure 4: The framework of the collaborative infrastructure. Collaboration must be established between units within the organization as well as between partners and customers outside the organization. Access to equipment and the visibility of strategic management are important at all levels of the structure [17].

architecture of human-robot systems including control, safety, and interface components [22]; security system frameworks to ensure the protection of data interaction during manufacturing [23]; and the exploration of approaches moving towards a cloud manufacturing ecosystem [24].

In summary, the requirements of horizontal integration are as follows:

(1) A need for tools that can optimize the total supply chain

(2) A need for tools that can simultaneously optimize different tasks such as scheduling, maintenance, production, and logistics

2.2. Features of Industry 4.0 Solutions: Further Requirements for the Development of Optimization Algorithms. In order to explore additional requirements, in this section, the features and further requirements of Industry 4.0 are presented. Industry 4.0 has clear, well-defined design principles that can refine its image. The requirements are defined based on the following four tenets [25]: interconnection, decentralized decision, information transparency, and technical assistance. These four design principles led to rapid development; moreover, the aforementioned principles were insufficient to fully characterize the aims and trends that have emerged. New properties have appeared that could provide a more detailed characterization of the design [26]. This phenomenon occurred again, and a more detailed description was necessary. Additional design considerations were identified as follows [27, 28]. The principles of the solutions and their relation to optimization are summarized in Table 2; moreover, examples are also given.

Horizontal and vertical integration can facilitate further optimization opportunities such as the Lean Six Sigma Tools [29]. They help to update the classical Manufacturing Execution System (MES) applications to monolithic Manufacturing Operations Management (MOM) [30], which forms the basis of up-to-the-mark cyber-physical systems (CPS) [31]. Horizontal and vertical integration can be supported by digital twin solutions. A publication reported the application of a digital twin, which provides service-based and real-time enabled infrastructure for vertical and horizontal integration [32]. Another paper introduced a viewpoint-oriented approach to identify functional components that facilitate decision-making within a manufacturing ecosystem [33].

In summary, Industry 4.0 is a more complex and connected concept than previous ones. It consists of many aspects that sometimes conflict with each other in practice. It only remains workable if the integration (vertical and horizontal) is an essential part of its implementation. Partly because of integration and partly because of the many data sources, a large amount of data are generated every second that need to be stored, processed, and aggregated. That is why Big Data is one of the key services in Industry 4.0. Decentralization is important in this concept, but the many services of Industry 4.0, e.g., data storage and decision-making, raise the following question: Is a decentralized, centralized, or mixed strategy the best choice in the given situation? The place/layer where decisions are made depends partly on feasibility and partly on the aforementioned strategy. This level of complexity is hardly or completely unmanageable. Therefore, the following question arises: Is the goal to identify the optimal operation or to search and maintain efficient and reliable operation? If the problem can

be broken down into smaller parts, they can already be simulated, and a multilevel simulation can be used for several overlapping subtasks. In addition to the traditional tasks, the simulations should also provide an answer to how the industrial process will be sustainable and fit into the circular economy (CE) because this viewpoint will become increasingly stronger in everyday life, so the companies have to account for it.

The name column in Table 2 shows the requirements and characteristics of Industry 4.0 except optimization. A description is given for every property that presents more information about it. Their effects on optimization are accumulated for each element based on our experience. If the column contains more "+" signs, the expected effect of that element on optimization will be greater. Every row shows an example of the relationship between optimization, Industry 4.0, and the given row in the given column. It can be clearly seen that our expectation is the following: the greatest effect is exerted by real-time capability, but corporate social responsibility, sustainability, flexibility, smart factories, virtualization, decentralization, modularity, interoperability, and smart products also affect it significantly. Furthermore, it is also worth noting that all requirements and characteristics of Industry 4.0 have an effect in this context. They are quantitatively analyzed in the following sections to determine whether our expectations are true or not.

3. Quantitative Analysis of the Related Publications

In this section, a quantitative analysis is presented concerning the available literature based on keywords, namely, Industry 4.0 and optimization. Such an analysis is very appropriate because the structure and topicality of the examined papers can be revealed. The publications over time show an ascending trend, and it can be observed that 80% of the papers were published between 2018 and 2020 (until October). The distribution of publications is shown in Figure 5.

The articles were published in many different scientific journals. It can be concluded that plenty of journals welcome papers on the topic of Industry 4.0. The found studies were published by more than 400 different journals or conference events. 25 journals that published the most studies are summarized in Table 3, and their distribution over the years is also included.

The study of keywords and their connections is also a practical aspect of the analysis. Keywords hide profitable information about the frameworks, methods, applications, or enabling technologies of the examined topic. The keywords that occur most frequently are gathered to show the importance weight of individual Industry 4.0 items, and their numbers of occurrences are visualized in Figure 6.

The linkage between Industry 4.0 and other keywords can be illustrated well separately using clustering algorithms. The connectivity of the keywords was obtained by VOSviewer software. The method gathers the keywords that occur together. Its hyperparameter is the minimum number of occurrences which was set at three. The results are illustrated in Figure 7.

The derived framework of keywords is separated according to fields of application, namely, wireless sensor network, smart grid, availability, and agriculture. An algorithm part is also present as well as genetic and evolutionary algorithms in addition to particle swarm optimization. Block scheduling also plays an important role and it is connected to the supply chain, modelling as well as cloud and lean manufacturing. Energy optimization is also an essential element of the network and is related to digital factories, sustainable manufacturing, and multiobjective optimization. The Internet of Things as the number one enabling technology of Industry 4.0 is also often referred to as a keyword and linked to cyber-physical systems, machine learning, modelling, and smart factories.

The integration and features of Industry 4.0 were introduced in Section 2. Their enabling and characterizing items can be found in the keywords. Vertical integration is characterized by the unification of company manufacturing levels for efficient production. Typical keywords appear, namely, Internet of Things, cyber-physical systems, smart manufacturing, energy optimization, control, monitoring, scheduling, productivity, and process optimization. Horizontal integration typically aims to create supply chain networks. As a result, the life cycle of the product can be traced. Specific keywords consist of supply chain, sustainability, energy optimization, business process optimization, cloud manufacturing, resource efficiency, and smart manufacturing networks.

A remarkably infrequent keyword is the operator, whose role is diminishing as a consequence of the development of Industry 4.0.

As has been seen previously in Section 2, Table 2 includes our expectations about the effects of the elements of Industry 4.0 on optimization. The data uncovered during the research are shown in Figure 8, which confirm the previous assumptions. Currently, real-time capability is the hottest topic in this context. A large number of works also deal with corporate social responsibility, flexibility, and smart factories. However, it is important to note that active research is underway in all the remaining areas.

The key message of the figure is that multiple goals usually need to be met simultaneously in the case of Industry 4.0, and the role of corporate social responsibility has continually increased. It is important to note that more and more goals can be managed in parallel as technology advances, which is increasingly opening the door to the wider usability of multiobjective optimizations.

These principles can be further discussed based on the related Industry 4.0 approaches. The concepts determined are associated with the given requirements. The concepts of the given requirements are defined, and to this end, connected keywords are included in parentheses in Table 4 based on Figure 7. The purpose of this table is to show which concepts should be used to achieve the different requirements and how it is possible to find cross-conceptions based on these keywords. For example, if autonomy should be achieved, one of the key concepts is machine

Table 2: Effect of the requirements and characteristics of Industry 4.0 on optimization.

Name	Description	Effect on optimization	Examples
Vertical integration	It integrates all logical layers within an organization, starting from the production floor to the business layer.	+	[34, 35]
Horizontal integration	It integrates data transfer across multiple production facilities or even the entire supply chain. A horizontally integrated company concentrates on the kinds of activities that are closely related to its competencies; moreover, it builds partnerships to support the end-to-end value chain.	+	[36, 37]
End-to-end engineering	It describes a process that takes a system or service from beginning to end and delivers a complete functional solution, usually without needing to obtain anything from a third party.	+	[38, 39]
Smart factory	It is a highly digitized and connected production facility that relies on smart manufacturing where the final goal is to organize the production facilities and logistics systems without human intervention.	++++	[40, 41]
Product personalization	It is a process of delivering customized goods and services to the customers as per their needs and desires.	++	[42]
Virtualization	It is the creation of a virtual—rather than actual—version of something, such as an operating system or production process.	++++	[43, 44]
Decentralization	It is a structure, where the activities of an organization, particularly those regarding planning and decision-making, are distributed or delegated away from a centralized infrastructure.	+++	[45, 46]
Flexibility	It is the ability to react to changes within a predetermined scope of requirements (corridor of action) time- and cost-effectively.	+++++	[47, 48]
Corporate social responsibility (CSR) and sustainability	It is a self-regulating business model that helps a company to be socially accountable to itself, its stakeholders, and the public. Sustainability focuses on meeting the needs of the present without compromising the ability of future generations to meet their needs.	++++++	[49, 50]
Real-time capability	It is the ability of a device or system to collect and analyze data as well as to respond instantaneously to a command, event, or input.	++++++++	[51, 52]
Modularity	It determines which independent and interlocking subsystems build the system based on their functionality.	+++	[53, 54]
Interoperability	It is the ability of systems to exchange information of unambiguous meaning.	+++	[55, 56]
Smart product	It is a data processing object, which has several interactive functions. It combines the physical and software interfaces; moreover, the usage of a smart product is interactive as well as requires also some cognitive work by the user.	+++	[57, 58]
Autonomy	It is the capacity to make an informed, uncoerced decision. Autonomous organizations, institutions, machines, or systems are independent or self-governing.	++	[59, 60]
Agility	It is defined as the ability of firms to sense environmental change and respond readily.	++	[61, 62]
Service orientation	It offers a service (of cyber-physical systems, humans, or smart factories) via the Internet of Services.	++	[63, 64]

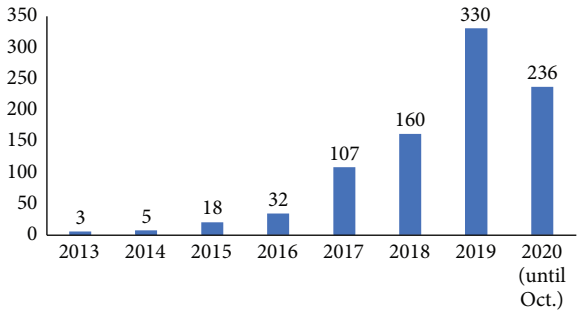


Figure 5: Number of publications per year.

Table 3: Articles by journal and period. The 25 most frequently publishing journals can be considered as well as the number of published papers each year.

Journal	Articles published per year								Total
	2013	2014	2015	2016	2017	2018	2019	2020*	
Procedia Manufacturing	0	0	1	0	11	6	11	9	38
Procedia Computer Science	0	0	0	0	0	0	17	15	32
IOP Conference Series: Materials Science and Engineering	0	0	0	0	1	2	23	4	30
IFAC-PapersOnLine	0	0	1	5	2	3	19	0	30
Procedia CIRP	0	0	0	1	6	4	8	5	24
Advances in Intelligent Systems and Computing	0	0	0	0	1	7	6	9	23
IEEE Access	0	0	0	1	6	0	5	5	17
IFIP Advances in Information and Communication Technology	1	0	0	0	5	1	4	5	16
IEEE ICA-ACCA 2018-IEEE International Conference on Automation/23rd Congress of the Chilean Association of Automatic Control: Towards an Industry 4.0-Proceedings	0	0	0	0	0	0	15	0	15
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	0	0	1	1	0	2	5	3	12
Journal of Physics: Conference Series	0	0	0	0	0	1	8	2	11
Applied Sciences (Switzerland)	0	0	0	0	0	0	4	7	11
Proceedings of SPIE-The International Society for Optical Engineering	0	0	0	0	1	1	5	2	9
Lecture Notes in Mechanical Engineering	0	0	0	0	0	2	0	7	9
International Journal of Advanced Manufacturing Technology	0	0	0	0	0	4	2	3	9
ZWF Zeitschrift fuer Wirtschaftlichen Fabrikbetrieb	1	1	3	1	2	0	1	0	9
Proceedings of the International Conference on Industrial Engineering and Operations Management	0	0	0	1	1	3	2	1	8
Sustainability (Switzerland)	0	0	0	0	0	1	2	5	8
Sensors (Switzerland)	0	0	0	0	0	0	1	6	7
Communications in Computer and Information Science	0	0	0	0	2	0	3	1	6
IEEE International Conference on Industrial Engineering and Engineering Management	0	0	0	0	0	2	4	0	6
Computer Aided Chemical Engineering	0	0	0	1	2	2	1	0	6
Proceedings of the Summer School Francesco Turco	0	0	0	1	2	2	1	0	6
2019 IEEE International Workshop on Metrology for Industry 4.0 and IoT, MetroInd 4.0 and IoT 2019-Proceedings	0	0	0	0	0	0	6	0	6
International Journal of Production Research	0	0	0	1	0	1	1	3	6

*Until October.

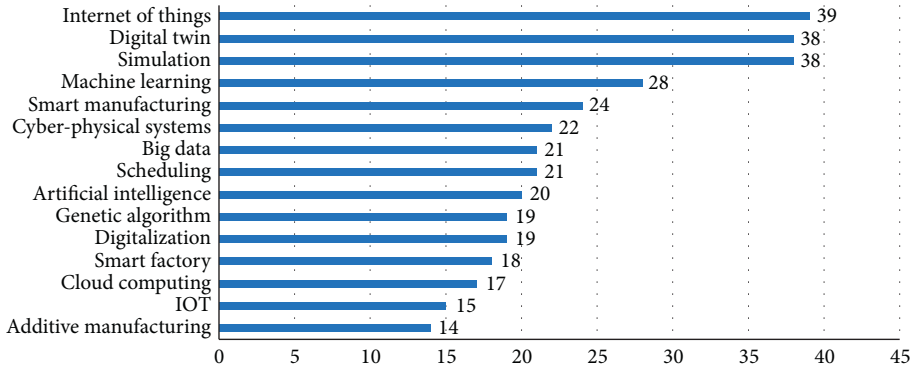


Figure 6: Frequency of the most commonly occurring keywords.

learning. Efficient learning is greatly supported by the Internet of Things as well as artificial intelligence. Data analysis, process optimization, and predictive maintenance can also be applied.

900 relevant papers can be found in the literature from 2013 to the present concerning the fields of Industry 4.0 and optimization. Table 5 shows the frequencies of occurrence of the major optimization areas in Industry 4.0, where the optimization tasks and their descriptions as well as some illustrative examples are given.

Industry 4.0 is a global and versatile topic, as revealed by the keyword analysis. The distribution of publications over the years has shown that much research aims to examine this subject. Plenty of journals are also available for reporting new solutions.

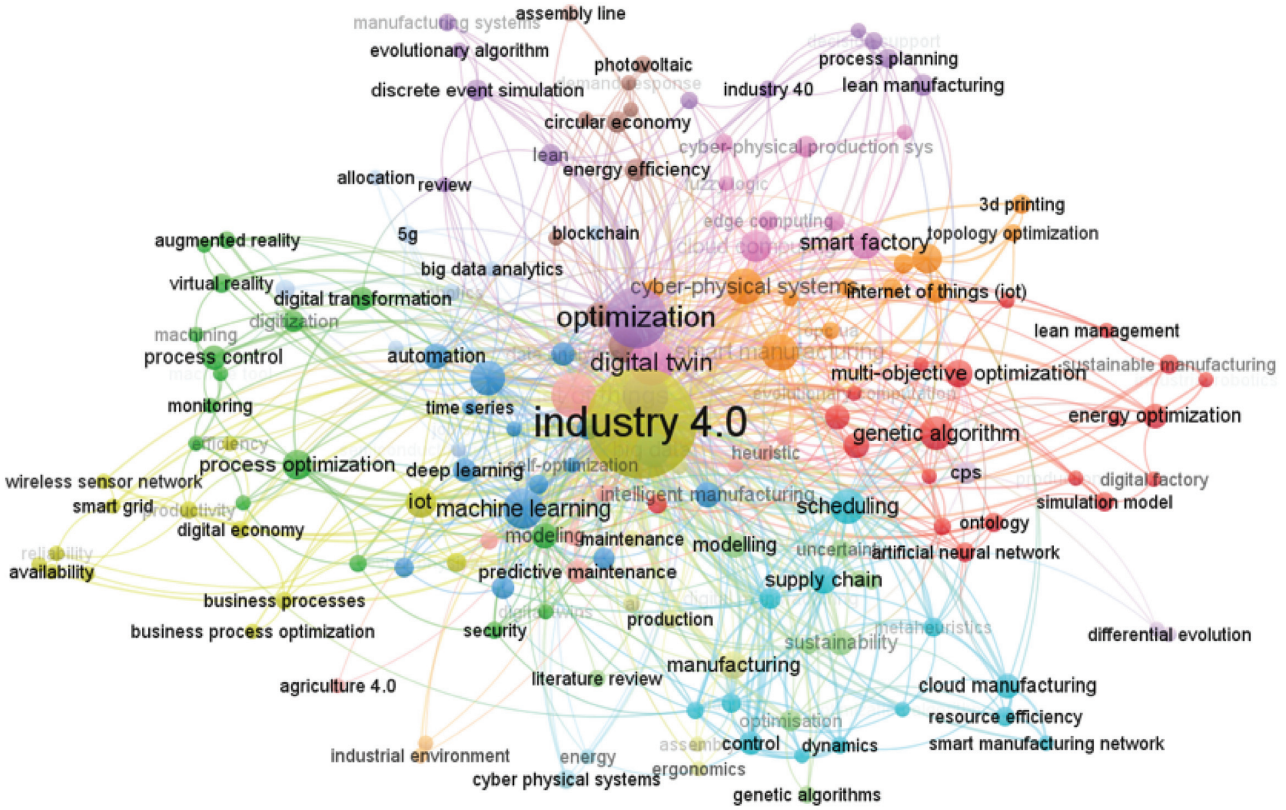


Figure 7: The linkage of keywords.

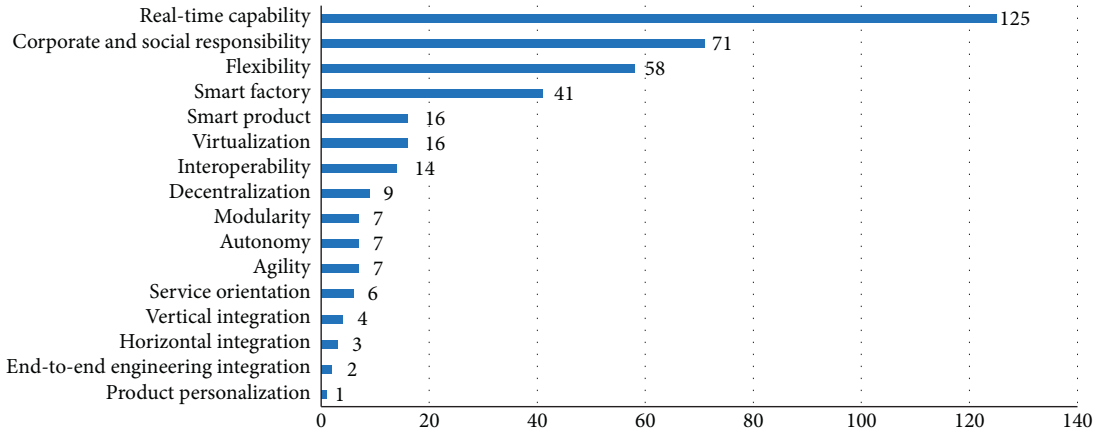


Figure 8: Linking optimization to the requirements and characteristics of Industry 4.0.

4. Analysis of Industry 4.0-Related Optimization Tasks and Algorithms

In this section, the Industry 4.0-related optimization tasks and algorithms are discussed. Typical optimization problems are generally formalized, their complexity and metric systems are determined, and the impact of changes on the development of these algorithms inducted by Industry 4.0-driven solutions is analyzed in Section 4.1. The characteristics of optimization tasks and the requirements of Industry 4.0 solutions are presented in Section 4.2. Future research areas are identified based on current trends in Section 4.3.

4.1. Formalization of the Optimization Problems. The nature of an optimization problem is determined by its constraints and objective function. The standard form of an optimization problem is as follows:

Table 4: Network analysis of keywords to show the concepts related to the different requirements as well as the keywords related to the concept in parentheses.

Requirement	Related concept (connected keyword)
Vertical integration	Internet of Things (machine learning, digital manufacturing, intelligent manufacturing, smart factory, lean management, process control, wireless sensor network, smart grid); business process (digitalization, Internet of Things, cyber-physical system, digital economy); process optimization (digital twin, machine learning, supply chain, digital transformation, anomaly detection); productivity (process optimization, automation, planning, efficiency, process automation); reliability (availability, smart grid, productivity, dependability); ontology (smart manufacturing, intelligent manufacturing, artificial neural network)
Horizontal integration	Lean management (Internet of Things, digitalization, sustainable manufacturing); ontology; smart city (Internet of Things, machine learning, cyber-physical system); security (machine learning, predictive maintenance, modeling, digital twin, anomaly detection); reinforcement learning (artificial intelligence, internet of things, deep learning, predictive maintenance, condition monitoring); resource efficiency (cloud manufacturing, sustainability, smart manufacturing network, additive manufacturing)
End-to-end engineering	Sustainable manufacturing (energy optimization, lean management, artificial intelligence, Internet of Things); cyber-physical system (digital twin, smart factory, flow shop scheduling); productivity; logistics (Internet of Things, intelligent manufacturing, digital twin, genetic algorithm, smart manufacturing)
Smart factories	Digital twin (cyber-physical system, smart factory); smart manufacturing (Internet of Things, digital transformation, data analytics, digital twin, supply chain, cyber-physical system, logistics); cyber-physical system; artificial intelligence (machine learning, Internet of Things, digitalization, digital twin, discrete event simulation); Internet of Things
Product personalization	Supply chain; job shop scheduling (automation, reinforcement learning, multi-agent systems); logistics (Internet of Things, intelligent manufacturing, digital twin, genetic algorithm, smart manufacturing)
Virtualization	Digital twin (cyber-physical systems, smart factory); virtual reality (digital transformation, digital twin, key performance indicator, machining)
Decentralization	Supply chain (scheduling, smart manufacturing, control, planning, digital transformation); blockchain (big data analytics, additive manufacturing, digital twin, Internet of Things); cloud computing (Internet of Things, genetic algorithm, evolutionary algorithm, digitalization)
Flexibility	Digital transformation (virtual reality, artificial intelligence, smart manufacturing, supply chain); assembly line (optimization, digital twin)
Corporate social responsibility	Optimization (machine learning, scheduling, energy efficiency); monitoring (process control, optimization, Internet of Things); efficiency (lean, digital economy); sustainability; smart grid (wireless sensor network, smart grid, reliability, efficiency); digital economy (business process, efficiency); resource efficiency (additive manufacturing, cloud manufacturing, sustainability, smart manufacturing network); circular economy (demand response, modelling); energy (control, modelling, cyber-physical system); energy optimization (digitalization, scheduling, digital factory, sustainable manufacturing); lean manufacturing (process planning, additive manufacturing, scheduling); sustainability (machine learning, optimization, modelling, resource efficiency)
Real-time capability	Monitoring; cyber-physical system (digital twin, smart factory, flow shop scheduling); Internet of Things; digital factory (energy optimization, simulation model); flow shop scheduling (genetic algorithm, particle swarm optimization, cyber-physical system, multi-agent system); lean management (Internet of Things, sustainable manufacturing); job shop scheduling; self-optimization (artificial intelligence, reinforcement learning, intelligent manufacturing); planning (productivity, manufacturing, scheduling, metaheuristic, control); scheduling (genetic algorithm, multi-objective optimization, cloud manufacturing, modelling, cyber-physical system); data acquisition (cyber-physical system, Internet of Things, smart manufacturing); demand response (circular economy, data-driven)
Modularity	Cloud computing (intelligent manufacturing, Internet of Things, genetic algorithm, data analytics); logistics; edge computing (Internet of Things, smart manufacturing, cloud computing, data analytics)
Interoperability	Internet of Things; sustainability; cyber-physical systems; smart factory (Internet of Things, digital twin, cyber-physical system, smart manufacturing, evolutionary algorithms, process planning); wireless sensor network (Internet of Things, smart grid); smart grid (Internet of Things, reliability, efficiency, wireless sensor network)
Smart product	Smart manufacturing (Internet of Things, digital transformation, data analytics, digital twin, supply chain, cyber-physical system, logistics); smart factory (Internet of Things, digital twin, cyber-physical systems, smart manufacturing, digital manufacturing, artificial intelligence, modelling)
Autonomy	Self-optimization (artificial intelligence, intelligent manufacturing, reinforcement learning, digitalization); deep learning (artificial intelligence, advanced process control, data analytics, machine learning, reinforcement learning); machine learning (internet of things, artificial intelligence, data analytic, predictive maintenance, process optimization); automation (artificial intelligence, digital twin, cyber-physical systems, digital transformation, logistics, review, planning)
Service orientation	Lean manufacturing (scheduling, additive manufacturing, process planning)
Agility	Wireless sensor network; smart grid; availability (interaction, dependability, reliability); reliability (availability, smart grid, productivity, dependability)

Table 5: I40-relevant application areas of optimization algorithms.

Keyword	Description	Illustrative applications	Number of relevant papers
Control	It helps to achieve and increase a consistent, economical, and safe production level	Controller for mobile robot path tracking [65] A control strategy for smart energy charging [66]	340
Planning	It is a preparatory step before manufacturing which optimizes several sequences of operations	Automating production planning and control in manufacturing [67] Operation planning of renewable energy [68]	138
Scheduling	It gives the optimal allocation of the available resources in the given workflows	Line balancing and AGV scheduling [59] Machine scheduling [69]	137
Maintenance	It defines optimization tasks for preventing the failure of expensive manufacturing equipment	Digital twin-driven autonomous maintenance [60] Sensor-based maintenance policies [70]	91
Energy efficiency	It helps to reduce the energy consumption of the manufacturing processes based on optimization techniques	Optimization of greenhouse production process [71] Energy-efficient scheduling by collaboration between cyber-physical production and energy systems [72]	32
Topology	It gives an answer as to how to place some materials on a given surface to obtain the best structural performance	Topology optimization of an automotive dashboard [73] Manufacturing redesign of a piece of military aircraft equipment [74]	14
Planning and scheduling	It is an integrated optimization of manufacturing process planning and scheduling	Integrated planning and scheduling problem in parallel is solved by a metaheuristic approach [75] Cloud-based intelligent dynamic planning and scheduling system [76]	12
Lean management	It focuses on minimizing waste and maximizing productivity within manufacturing systems simultaneously	Process optimization by lean management [77]	9
Business process optimization	It can increase the efficiency of the production, delivery, and operational processes amongst others in industrial environments	Improving production processes [78]	5
Optimal control	It is an optimization of control processes and subsystems to support an efficient operation	Scheduling flexibly configurable jobs [52]	5
Resource optimization	It helps align available resources with the maximum efficiency of goals	Resource optimization of industrial processes [79]	2

$$\begin{aligned} &\underset{x}{\text{minimize}}\ f(\mathbf{x}) \\ &\qquad g_i(\mathbf{x}) \le 0,\ i = 1, \ldots, m, \\ &\text{subject to} \quad h_j(\mathbf{x}) = 0,\ j = 1, \ldots, l, \\ &\qquad \mathbf{x} \in \mathbf{R}, \end{aligned} \tag{1}$$

where $f, g_1, \ldots, g_m, h_1, \ldots, h_l$ are $\mathbf{R}^n \longrightarrow \mathbf{R}$ functions.

In the following description, the models are discussed in ascending order of their computational complexity.

If an optimization task does not contain constraints, then it is a so-called unconstrained optimization problem regardless of the nature of the objective function. These tasks can be handled with traditional tools of mathematical analysis.

Optimization problems that contain only linear constraints and a linear objective function are called linear programming (LP) problems:

$$\begin{aligned} &\text{minimize}\ c_0 + \sum_{j=1}^{n} c_j x_j \\ &\text{subject to} \quad L_i \le \sum_{j=1}^{n} a_j^i x_j \le U_i,\ i = 1, \ldots, m, \\ &\qquad l_j \le x_j \le u_j,\ j = 1, \ldots, n. \end{aligned} \tag{2}$$

The existence of effective LP solvers is very important in optimization because they can solve a wide range of optimization problems and are often hidden engines of other algorithms. For example, lots of nonlinear models can be transformed into an equivalent LP model or solved by solving a sequence of linear models; moreover, integer programming solvers usually solve lots of LP models. The two most used solvers are the simplex method and the interior-point method.

Typical Industry 4.0-focused linear optimization problems are either product-mix problems, where the goal is to determine the optimally produced quantities of product, or transportation problems consisting of the optimal logistics in an efficient way.

The next group of optimization problems is quadratic programming (QP) where, under linear constraints, the objective function is as follows: minimize $(1/2)\mathbf{x}^T Q\mathbf{x} + c^T\mathbf{x}$, where Q denotes an $n \times n$-dimensional real symmetric matrix, c stands for a real-valued, i-dimensional vector, and the decision variables are represented by the n-dimensional vector $\mathbf{x}$. A variety of methods can be used to handle the problem, e.g., the extended simplex algorithm and interior-point method. Typical QP problems are regression-based optimization problems that optimize supply and demand response, where the equilibrium price and quantity are modelled. Many QP problems are highlighted in paper [80].

The most complex optimization tasks are the group of nonlinear optimization problems (NLP), where the objective function is a nonlinear function of the decision variables and the constraints can be linear as well as nonlinear expressions. Typical Industry 4.0-focused nonlinear optimization problems are as follows: optimal control, where a control plan for a dynamical system over a period of time is optimized, and predictive maintenance, where the aim is to estimate how often system maintenance should be performed, but priority areas are energy efficiency and topology optimization, all of which effectively support the improvement of the efficiency of complex industrial systems.

A new version of the aforementioned optimization problems is obtained when the decision variables are not continuous but discrete. In the general case, this means an exponential number of steps are required to solve the tasks. Therefore, in the case of linear programming tasks, integer programming (IP) and mixed-integer linear programming (MILP), if both integer and continuous variables can occur between the decision variables, can be referred to. The standard form of the problem is as follows:

$$\begin{aligned} &\text{minimize } c^T\mathbf{x} + h^T\mathbf{y} \\ &\qquad\qquad A\mathbf{x} + D\mathbf{y} = b \\ &\text{subject to} \quad \mathbf{x}, \mathbf{y} \geq 0 \\ &\qquad\qquad \mathbf{x} \text{ integer} \left(\mathbf{x} \in \mathbf{Z}^n\right). \end{aligned} \tag{3}$$

Very often, the integer variables are bounded, i.e., $0 \leq \mathbf{x} \leq u$. To efficiently handle an optimization problem, a suitable methodology and mathematical solver need to be chosen to take into account the nature of the problem. An integer programming model of the assignment problem is an illustrative example here, where the jobs and workers are assigned in an optimal way. Similarly, scheduling problems can be described by (mixed) integer models that help optimize both manufacturing steps and business processes in the manufacturing environment.

The multiobjective optimization model is developed when more than one objective function is optimized simultaneously. In the case of nontrivial models, it can be said that no single solution is optimal for all goals, since the objective functions work against each other. Basically, two approaches are available to solve multiobjective optimization models: traditional methods using single-objective functions and optimization algorithms based on a Pareto front. The nature of the objective functions and the constraint conditions determine the complexity of as well as difficulty in solving the problem. A review of multiobjective optimization frameworks, models, and algorithms is summarized in [81]. The standard form of the multiobjective problem is as follows:

$$\begin{aligned} &\text{minimize} \mathbf{z} = \{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\} \\ &\qquad\qquad g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, s_1 \\ &\text{subject to} \quad h_j(\mathbf{x}) = 0,\ j = 1, \ldots, s_2 \\ &\qquad\qquad l_k \leq x_k \leq u_k,\ k = 1, \ldots, n. \end{aligned} \tag{4}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denotes an n-dimensional vector of decision variables and $\mathbf{z} = (z_1, z_2, \ldots, z_m)$ stands for an m-dimensional vector of objective functions. Function g_i represents the inequality constraint i and function h_j yields the equality constraint j, where the number of these constraints are s_1 and s_2, respectively. The lower and upper bounds on the decision variable x_k are represented by l_k and u_k. Regardless of the mathematical model, each optimization problem can be converted into a multiobjective programming problem if further objective functions are added to the set of constraints (for example, should not only cost but also time and reliability be optimized).

Modelling and handling uncertainties cannot be avoided in Industry 4.0-focused optimization tasks. Different modelling strategies exist in terms of optimization that takes into account uncertainties: one approach is stochastic programming, where the standard two-stage paradigm can be applied. Here, the goal is to choose the first-stage variables in such a way that the sum of the overall cost during the first stage and the expected value of the second stage is minimal. The stochastic programming model can be a linear, integer, or nonlinear problem, depending on the nature of the given task. The standard formulation of the two-stage stochastic linear program is as follows:

$$\begin{aligned} &\text{minimize} c^T\mathbf{x} + E_\xi[Q(\mathbf{x}, \xi)] \\ &\text{subject to} \quad A\mathbf{x} = b,\ \mathbf{x} \geq 0 \\ &\text{where} \quad Q(\mathbf{x}, \xi) = \text{minimize}\left\{q^T\mathbf{y} | T\mathbf{x} + W\mathbf{y} = h,\ \mathbf{y} \geq 0\right\}, \end{aligned} \tag{5}$$

where vectors $\mathbf{x}$ and $\mathbf{y}$ represent the first- and second-stage decision variables. The elements T, W, h, and q of the second stage can be random, and thus, $\xi = (q, h, T, W)$ is a random vector. Note that matrices T and W are called the technology and recourse matrices, respectively. By transforming the objective function, a stochastic programming model can also be considered as robust, where the objective function also includes the measure of variability and the risk tolerance of the modeller. In addition, if the reliability of the given system is highlighted and uncertainty in the parameters of the constraints is assumed, then the probabilistic programming approach can provide the right model representation.

Another approach to modelling optimization problems that takes into account uncertainties is fuzzy mathematical programming. The fundamental difference between stochastic and fuzzy optimizations is in the description of uncertainty since the membership function represents the different realizations of uncertainty. Based on the above, the parameters, constraints, and the objective function can be represented by fuzzy numbers, fuzzy sets, and the fuzzy goal, respectively. In such cases, the flexible programming and possibilistic programming strategies can provide appropriate solutions. Finally, it is also worth mentioning solutions based on stochastic dynamic programming, since uncertainty can be an integral part of the technique. The solution assumes that a given time point is entirely determined by its history. Therefore, first, the tail subproblems are managed, from which the solution to the original problem can be derived. To determine the solution, a large number of computation steps and time requirements must be used in the calculation, which increases exponentially as a function of the number of variables in the model. In recent years, a number of papers have been published in the fields of stochastic programming, fuzzy programming, and stochastic dynamic programming related to Industry 4.0 applications, e.g., production planning, scheduling, capacity expansion, energy investment as well as the design and optimization of complex chemical engineering systems. Sahinidis presents a state-of-the-art optimization that takes into account uncertainty in [82].

Many optimization problems can be represented by networks such as Petri nets [83], P-graphs [84], and State-Task Networks [85]. These approaches can visualize the problem to aid understanding. Efficient solution algorithms usually exist for network type problems which can exploit the structure of the problem. Unlike general methods, these algorithms are specialized, i.e., cannot solve other types of problems.

Two groups of solvers can be used to solve MIP models, namely, exact optimization algorithms and approximate algorithms. Exact algorithms can always generate the optimal solution of the model, but the size of the solvable models is insufficient for industrial needs. Approximate algorithms are faster, so they can solve bigger problems, but usually, the quality of the solution cannot be guaranteed. Numerous solution methods are available, but the aim of this paper is not to review all of these traditional algorithms.

The types of most typical optimization problems are given in Table 6.

4.2. Characteristics of Optimization Tasks: Requirements of Industry 4.0 Solutions. Classical optimization models have typical characteristics. These properties are summarized in Table 7 based on [118]. The table lists the key features of an optimization task that are critical to the solution.

The size of a system affects the size of its model, and the size of a model drastically affects its solution time. Determining the optimal solution of a large problem is difficult as the computational time may be impractical. Horizontal and vertical integration increases the size of a suitable model.

Every manufacturing system has some kinds of natural modular structure where tight or loose connections are found between the modules. Modules can have their own optimization models, and their connections can be modelled using additional constraints. This modularity can help to solve a model using multistage optimization techniques; moreover, the models of modules can be solved in parallel. Multistage optimization techniques are usually faster than traditional ones but cannot guarantee to calculate the global optimum.

The complexity of words has multiple meanings in terms of optimization. The Oxford English Dictionary defines complexity as, "The state or quality of being intricate, complicated or complex." This meaning can be used as a property of a manufacturing system (system complexity). Wiendahl and Scholtissek [119] divide the definition of complexity in industrial manufacturing into the complexity of the products and complexity in production. Bozarth et al. [120] introduced detail complexity as the distinct number of components of the system and dynamic complexity as the unpredictability of a system's response. Deshmukh et al. [121] used the terms static complexity which can be viewed as "a function of the structure of the system, connective patterns, variety of components, and the strengths of interactions" and dynamic complexity which is "concerned with unpredictability in the behavior of the system over a time period." A mathematical model also can be complex in terms of system complexity, i.e., it can consist of many variables and constraints; moreover, the objective function and/or the constraints can be nonlinear. Usually, a complex system results in a complex model.

The computational complexity (also referred to as the time complexity) means a classification of algorithms and problems. For an algorithm, the classification means the number of steps needed to solve a problem according to the size of the input. From this point of view, polynomial algorithms can be distinguished from nonpolynomial ones since polynomial algorithms are faster. The complexity of a problem is the complexity of the best algorithm that can solve it. As a result, P (polynomial) problems can be distinguished from NP (nonpolynomial) ones. If a problem belongs to the NP class, a polynomial algorithm has yet to be found which can solve it. By increasing the size of an NP problem, its solution time is increased in a nonpolynomial way (e.g., exponential, factorial, and double exponential). Ausiello et al. [122] identified more than 200 NP optimization problems. In manufacturing, many well-known NP problems can be found, e.g., scheduling and process network synthesis. If the computational complexity of a model is high, then the solution time can be impractical for problems on an industrial scale.

An exact algorithm can generate an optimal (or near-optimal) solution for the model and an optimal solution for the model which is not necessarily an optimal solution to the problem. The quality of the solution depends on the quality of the model, i.e., only a good model results in a good solution. In the case of approximate methods, the quality of the solution not only depends on the quality of the model but also on the methods because these algorithms usually cannot guarantee optimality.

Table 6: Types of optimization problems.

Optimization models	Model type	Example
Stochastic programming	Uncertainty optimization	[69, 86–88]
Robust optimization	Uncertainty optimization	[89–91]
Global optimization	Deterministic, continuous, unconstrained/constrained optimization	[92–94]
Mixed-integer nonlinear programming	Deterministic, continuous, constrained, nonlinear programming	[95–97]
Network optimization	Deterministic, continuous, constrained optimization	[98, 99]
Derivative-free optimization	Deterministic, continuous, constrained, bound constrained optimization	[100–102]
Quadratic programming	Deterministic, continuous, constrained, linearly constrained optimization	[103–105]
Linear programming	Deterministic, continuous, constrained, linearly constrained optimization	[106–108]
Integer linear programming	Deterministic, discrete	[109–111]
Combinatorial optimization	Deterministic, discrete	[112–114]
Multiobjective optimization	—	[115–117]

Table 7: Typical characteristics of classical optimization models.

Property	Description	Trend
Size	The number of decision variables, parameters, and constraints	Increasing continuously because the vertical and horizontal integration increases the size of the manufacturing system (supply chain) to be modelled
Modularity	High level of modularity can help to solve the model efficiently	Important because the production systems are modular and have hierarchic structure; moreover, the modules may be solved independently decreasing the necessary solution time, and they can be solved parallel
Complexity	Complexity and nonlinearity may affect the solution time and precision dramatically	Increasing because only a complex and detailed model can describe a complex system appropriately
Time scale/ adaptability	Optimization techniques often require a relatively long time to give the solution	Important because the computational time can be critical for short-term problems especially for real-time problems; moreover, the flexibility of manufacturing systems is an expectation
Solution quality	A measure of how far a solution is at most from an optimal solution is obtained	Important for long-term problems, but for short-term problems the computational time is more critical

4.3. Future Research and Development Directions. As is mentioned above, the solution time is very critical in industry, especially for real-time problems. For NP problems, no algorithms are known, which can quickly generate their optimal solutions. To obtain a good but not necessarily optimal solution, fast heuristic approaches are needed. Heuristics should result in simpler models where some less important details are neglected or in heuristics-based algorithms, e.g., genetic algorithm, ant colony algorithm, tabu search, and artificial intelligence-based approaches.

One of the reasons for the existence of real-time problems is the uncertainty of systems which arises from unexpected tasks and events, e.g., a breakdown of a machine which is in need of maintenance [123]. Furthermore, the environment in which the production system has to work is uncertain and constantly changing. This property originates from changes in customer demands, product design, and processing technology. Handling these uncertainties is a challenge in many fields [124, 125]. Obviously, it becomes ambiguous which solution can be considered as optimal, e.g., the most robust one and the one expected to generate the highest profit. In this area, two main directions of research exist. In the first, the solution must be determined a priori and cannot be modified after the realization of an uncertain event. In the second, a plan exists, and when an uncertain event occurs, a reoptimization must be performed. Unlike in the previous case, the optimization is carried out when the system is already running; therefore, usually a very limited amount of time is available to deliver the solution. As a result, heuristics are often favored for this purpose because for real-time problems, a feasible solution should be sufficient which does not cause big changes compared to the original plan. Naturally, many developments do not fit into these two categories, e.g., in two-stage methods, the first stage defines an initial plan, but some decisions are later altered in the second stage during its operation.

4.3.1. Multistage Optimization. Multistage optimization is very frequent in Industry 4.0 because (horizontal and vertical) integration yields large-scale models which cannot be solved in a reasonable period of time. In high-level stages, only the "important" decisions are made and the "less important" ones are relaxed, while in lower-level stages, the "important" decisions come from the higher levels. For example, the ISA-95 standard defines 4 stages, namely, production planning, production scheduling, operating instructions, and process optimization (see Figure 2). Production planning defines which products will be manufactured; production scheduling provides the timetable of the manufacturing of the selected products, and the last two stages deal with optimizing the operations. The decentralization requirement of Industry 4.0 also points to multistage optimization because it also results in the decentralization of

decision-making. For example, in the case of scheduling problems, Rossit et al. [126] showed that this decentralization is one of the main research directions.

Rolling horizon decision-making is a good choice for optimal decisions in a stochastic, dynamically changing environment. It is the task of the decision-maker to make a decision for a certain number of future periods in such a way that the optimal decisions made for the first period effectively support future plans. The second decision time period starts with a replanning periodicity step where previously made decisions can be revised and updated. This redesign procedure is repeated for every time period to refine and clarify future plans, i.e., rolling horizon systems iteratively solve a series of optimization problems over a shorter period. One of the most common applications of rolling horizon designs is connected to production planning problems in which decisions are made where demand is satisfied at a minimal cost. The rolling horizon planning problem has been classified by Sahin et al. [127] who focused on deterministic and stochastic demands, single and multiple planning layers, single and multiple manufacturing layers, and single and multiple items. Rolling horizon optimization problems mainly focus on optimizing the different parameters of the planning problem, e.g., the length of the planning horizon, stock level, and forecast errors. The integration of planning and scheduling in a rolling horizon is taken much less into account due to the complexity of the optimization tasks examined. Rolling operation poses some Industry 4.0-focused optimization tasks, e.g., production scheduling, supply chain planning, optimization of planning horizon parameters and lot-sizing and scheduling decisions, the investigation of the planning frequency in production systems, the impact of forecast errors in planning processes, or the examination of the lot-sizing, parallel machine lot-sizing, and scheduling problems.

4.3.2. Digital Twin. Cyber-physical systems often have digital twins as building elements. The cyber part of a digital twin is suitable for performing various calculations and simulations [128] and, moreover, can provide information about optimization and scheduling algorithms [129]. Knowledge-based digital twin solutions can also help optimize the production [130] as they can be used to support decisions based on simulations, predictions, and optimization before providing intelligent real-time control as is depicted in Figure 9.

Digital twin solutions have already been applied for ergonomic optimization [131], optimizing the behavior of the system during the design phase [132] and multirobot manufacturing cell optimization [47]. Process simulations are also possible [133]. The approach seems to have become a key enabler of further Industry 4.0 solutions and their optimizations in the future, e.g., autonomous maintenance [60]. There has never been a full digital twin application in the field of pharmaceutical manufacturing [134]. Furthermore, this technology has never been implemented in cloud platforms; however, there is a need to build connections between these Industry 4.0 solutions [135]. The support of optimal product design using digital twin solutions is also a serious challenge for future research [136].

4.3.3. Ontology-Based Knowledge Representation and Integration. Ontology is a formal model that uses mathematical logic to clarify and define things as well as their relations [137]. Since they can be described by mathematical models, it is possible to implement them into a computerized environment.

Ontology-based design is not yet widespread and, therefore, seems to be a future research direction. However, initiatives for cyber-physical systems already exist [138]. Ontologies have also been constructed to support the decision-making of agents, and therefore, the agility and flexibility requirements have been met too [139].

Based on XML, several ontology formats have been developed such as B2MML (Business to Manufacturing Markup Language) and AutomationML (Automation Markup Language). B2MML is based on the ISA-95 standard and aims to support the vertical integration process with appropriate data conversion solutions and the efficient design of information exchange [140]. Firstly, useful information to be exchanged has to be identified and aspects are commonly provided by process models [141]. Subsequently, a control system supervises and schedules the exchange of data. The workflow using B2MML can be easily presented for scheduling [142]. This cycle is illustrated in Figure 10. The production data must be collected first via B2MML which is the input of the scheduling algorithm. The output is shared with the given dispatching system and compared with the process model. If the process seems to be inefficient, the production parameters are changed.

AutomationML (Automation Markup Language) is an XML-based data format for exchanging information between modern automation engineering tools [143]. It provides a hierarchical description language for industrial systems that takes into consideration both structure and properties [144]. AutomationML is very popular, widely accepted, and supported by different companies as well as suppliers. By using AutomationML, the overall process becomes faster and independent of the knowledge and experience of design engineers [145]. The structure of AutomationML consists of individual devices called "instances." They can be classified hierarchically into groups represented as "system unit classes." The entire system is named an "instance hierarchy." The interconnections between units are called "internal links" [146]. AutomationML offers suitable data collection and preprocessing opportunities by being combined with model-driven engineering [147]. This integration can reduce manual effort and increase reusability. AutomationML can be extended by performance metrics and used for optimization [144].

5. Discussion and Conclusions

Industry 4.0 aims for comprehensive automation and integration to achieve the most efficient, optimal operation. Although countless studies have been published, no review

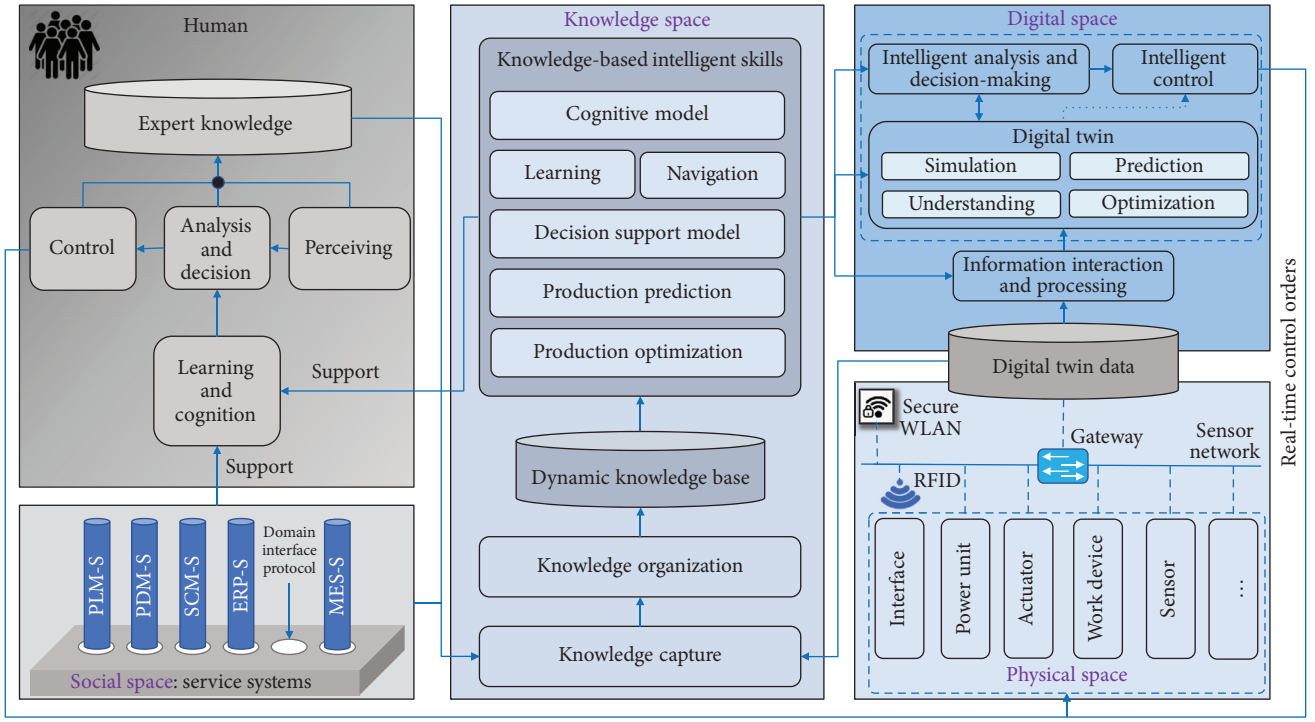


Figure 9: Architecture of knowledge-based digital twin solutions. The digital world and humans can collaboratively control the physical space and the knowledge space that supports it. The experiences are also saved and, therefore, captured.

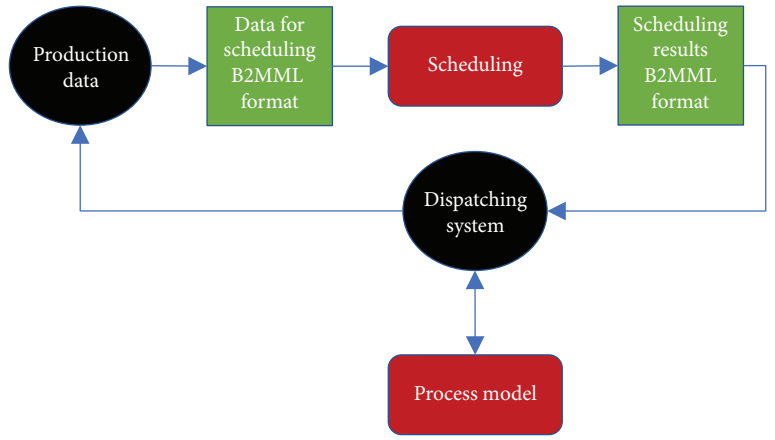


Figure 10: The workflow using B2MML for scheduling.

papers have been written about the optimization solutions of Industry 4.0. This paper provides an overview of the optimization methods and algorithms defined by the Industry 4.0 trend. The driving force behind these solutions is primarily integration, vertical and horizontal in particular.

Vertical integration aims to achieve the internal integration of enterprises from the production level to the business level. Planning scheduling and operation tasks are becoming increasingly dependent on each other, so the three aforementioned tasks need to be addressed simultaneously, and in an integrated way, moreover, real-time optimization plays an important role. Horizontal integration aims to organize the total supply chain management from the supplier to the customer, which needs to be thoroughly optimized. Tools must be developed for the simultaneous optimization of different global tasks such as scheduling, maintenance, production, and logistics. Horizontal integration also facilitates data exchange for analyzing processes, and its investigation can identify inefficient processes. A related concept known as collaborative manufacturing management has been presented. It is also clear that horizontal and vertical integration must be applied simultaneously to achieve the most efficient operation and also facilitate further optimization opportunities such as the implementation of lean tools.

The requirements of Industry 4.0 have been collected and discussed. These principles can provide guidance while developing new Industry 4.0 solutions. All of these

requirements have different effects on optimization and sometimes stand in conflict with each other in practice. However, solutions and concepts must be found to meet these requirements. Some examples were collected, which offer ideas for development. Keywords, which can define concepts, were also investigated by network analysis and help solve a given requirement properly based on the linkages. Such a table has also been presented which can determine research and development directions. Keyword analysis also revealed the connections between keywords and their structure. The derived framework is distributed for different areas like applications, algorithms, and enabling technologies.

After the discussion of integration methods and requirements, a quantitative analysis of the related publications was presented. The publications over time show an ascending trend and it can be observed that 80% of the papers were published by many journals between 2018 and 2020 (until October). The number of most frequently occurring keywords in published papers about given requirements was counted. Internet of Things, digital twin, and simulations were the keywords that most frequently occurred. Real-time capability, corporate social responsibility, and flexibility are the top 3 most researched topics among the requirements. Relevant application areas and their illustrative examples were also collected. Control, planning, and scheduling were the most popular topics.

A large number of optimization problems to be solved in industrial environments exist which, in many cases, require the application of complex solution methods. This paper has identified a range of key Industry 4.0 focus issues such as management, planning, scheduling, and maintenance, tasks of which all can be described as complex optimization problems through their interconnections. As has been pointed out, the starting point of optimization tasks is the continuous, integer, and mixed-integer linear programming tasks, on which the processes of the production environment can be modelled and examined more and more accurately. Given the uncertainty of modelling, incorporating the diversity of objective functions into the model and considering nonlinear relationships help the design to become increasingly accurate. Based on all of the aforementioned points, it is clear that the size of the mathematical models set up, their complexity, the modularity of the problem, and solution stability are key factors. All these have a decisive effect on the solution of the set optimization tasks, their accuracy, and their management using computer software. The difficulty in determining a solution is well illustrated by the fact that solving integer-value optimization tasks requires an exponential number of steps from mathematical solvers. For example, a scheduling task itself formulates a difficult NP task, which is integrated many times with other production and operational requirements; therefore, identifying the optimal solution is a growing challenge. All of these difficulties can be reduced by decomposing models as well as applying effective modelling and heuristics. With formal methods, the optimal solution can be validated with simulation tools; moreover, by using digital twins, the detailed efficiency-improving role can be seen as well.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the TKP2020-NKA-10 Project financed under the 2020-4.1.1-TKP2020 Thematic Excellence Programme by the National Research, Development and Innovation Fund of Hungary.

References

[1] C. Ierardi, L. Orihuela, and I. Jurado, "Distributed estimation techniques for cyber-physical systems: a systematic review," *Sensors*, vol. 19, no. 21, p. 4720, 2019.

[2] M. Dziubany, L. Creutz, S. Kopp, J. Schneider, A. Schmeink, and G. Dartmann, "Development of a cyber-physical system for an autonomous indoor transportation service," in *Proceedings of the 2020 9th Mediterranean Conference on Embedded Computing (MECO)*, pp. 1–5, IEEE, Budva, Montenegro, June 2020.

[3] B. A. Osei, N. A. Ragavan, and H. K. Mensah, "Prospects of the fourth industrial revolution for the hospitality industry: a literature review," *Journal of Hospitality and Tourism Technology*, vol. 11, no. 3, pp. 479–494, 2020.

[4] L. Monostori, B. Kádár, T. Bauernhansl et al., "Cyber-physical systems in manufacturing," *CIRP Annals*, vol. 65, no. 2, pp. 621–641, 2016.

[5] N. Sihag and K. S. Sangwan, "A systematic literature review on machine tool energy consumption," *Journal of Cleaner Production*, vol. 275, p. 123125, 2020.

[6] A. Angelopoulos, E. T. Michailidis, N. Nomikos et al., "Tackling faults in the industry 4.0 era—a survey of machine-learning solutions and key aspects," *Sensors (Switzerland)*, vol. 20, no. 1, p. 109, 2020.

[7] A. Goienetxea Uriarte, A. H. C. Ng, and M. Urenda Moris, "Bringing together lean and simulation: a comprehensive review," *International Journal of Production Research*, vol. 58, no. 1, pp. 87–117, 2020.

[8] J. Prinsloo, S. Sinha, and B. von Solms, "A review of industry 4.0 manufacturing process security risks," *Applied Sciences*, vol. 9, no. 23, p. 5105, 2019.

[9] D. Moher, L. Shamseer, M. Clarke et al., "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement," *Systematic Reviews*, vol. 4, no. 1, p. 1, 2015.

[10] F. S. Costa, S. M. Nassar, S. Gusmeroli et al., "FASTEN IIoT: an open real-time platform for vertical, horizontal and end-to-end integration," *Sensors*, vol. 20, no. 19, p. 5499, 2020.

[11] A. Seyedamir, B. R. Ferrer, and J. L. M. Lastra, "An ISA-95 based ontology for manufacturing systems knowledge description extended with semantic rules," in *Proceedings of the 2018 IEEE 16th International Conference on Industrial Informatics (INDIN)*, pp. 374–380, IEEE, Porto, Portugal, July 2018.

[12] B. Wally, C. Huemer, and A. Mazak, "Entwining plant engineering data and ERP information: vertical integration with AutomationML and ISA-95," in *Proceedings of the 2017 3rd International Conference on Control, Automation and Robotics (ICCAR)*, pp. 356–364, IEEE, Nagoya, Japan, April 2017.

[13] A. Semmar, N. Machkour, R. Boutaleb et al., "Modeling input data of control system of a mining production unit

based on ISA-95 approach," in *Proceedings of the International Conference on Smart Applications and Data Analysis*, pp. 47–55, Springer, Marrakesh, Morocco, June 2020.

[14] J. Sun, H. Yamamoto, and M. Matsui, "Horizontal integration management: an optimal switching model for parallel production system with multiple periods in smart supply chain environment," *International Journal of Production Economics*, vol. 221, p. 107475, 2020.

[15] J. Bechtold, A. Kern, C. Lauenstein, and L. Bernhofer, *Industry 4.0-The Capgemini Consulting View* Capgemini Consulting, Paris, France, November 2020, https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/07/capgemini-consulting-industrie-4.0_0_0.pdf.

[16] X. Xue, Y. M. Kou, S. F. Wang, and Z. Z. Liu, "Computational experiment research on the equalization-oriented service strategy in collaborative manufacturing," *IEEE Transactions on Services Computing*, vol. 11, no. 2, pp. 369–383, 2016.

[17] ARC Advisory Group, *Collaborative Manufacturing Management Strategies*, ARC Strategies, Sacramento, CA, USA, 2002.

[18] M. R. Firmansyah and Y. Amer, "A review of collaborative manufacturing network models," *International Journal of Materials, Mechanics and Manufacturing*, vol. 1, no. 1, pp. 6–12, 2013.

[19] M. Mladineo, S. Celar, L. Celent, and M. Crnjac, "Selecting manufacturing partners in push and pull-type smart collaborative networks," *Advanced Engineering Informatics*, vol. 38, pp. 291–305, 2018.

[20] S. Kuik and L. Diong, "A model-driven decision approach to collaborative planning and obsolescence for manufacturing operations," *Industrial Management & Data Systems*, vol. 119, no. 9, pp. 1926–1946, 2019.

[21] Y. Cheng, L. Bi, F. Tao, and P. Ji, "Hypernetwork-based manufacturing service scheduling for distributed and collaborative manufacturing operations towards smart manufacturing," *Journal of Intelligent Manufacturing*, vol. 31, no. 7, pp. 1707–1720, 2018.

[22] L. Pérez, S. Rodríguez-Jiménez, N. Rodríguez, R. Usamentiaga, D. F. García, and L. Wang, "Symbiotic human–robot collaborative approach for increased productivity and enhanced safety in the aerospace manufacturing industry," *The International Journal of Advanced Manufacturing Technology*, vol. 106, no. 3-4, pp. 851–863, 2020.

[23] K. Lu, Q. Han, G. Zhu, and B. Huang, "Research on the whole process security system framework of network collaborative manufacture," in *Proceedings of the 2020 Chinese Control and Decision Conference (CCDC)*, pp. 5530–5534, IEEE, Hefei, China, August 2020.

[24] P. Helo, Y. Hao, R. Toshev, and V. Boldosova, "Cloud manufacturing ecosystem analysis and design," *Robotics and Computer-Integrated Manufacturing*, vol. 67, p. 102050, 2021.

[25] W. de Paula Ferreira, F. Armellini, and L. A. De Santa-Eulalia, "Simulation in industry 4.0: a state-of-the-art review," *Computers & Industrial Engineering*, vol. 149, p. 106868, 2020.

[26] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business & Information Systems Engineering*, vol. 6, no. 4, pp. 239–242, 2014.

[27] M. Hermann, T. Pentek, and B. Otto, "Design principles for industrie 4.0 scenarios: a literature review," in *Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS)*, vol. 1, pp. 3–15, Koloa, HI, USA, January 2015.

[28] M. Hermann, T. Pentek, and B. Otto, "Design principles for industrie 4.0 scenarios," in *Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 3928–3937, Koloa, HI, USA, January 2016.

[29] A. Chiarini and M. Kumar, "Lean six sigma and industry 4.0 integration for operational excellence: evidence from Italian manufacturing companies," *Production Planning and Control*, pp. 1–18, 2020.

[30] M. Meyer-Hentschel, O. Lohse, S. Rao, and R. Lepratti, "Manufacturing operations management for smart manufacturing-a case study," in *Advances in Production Management Systems. The Path to Digital Transformation and Innovation of Production Management Systems*, pp. 91–98, Springer International Publishing, Cham, Switzerland, 2020.

[31] A. Ahmadi, A. H. Sodhro, C. Cherifi, V. Cheutet, and Y. Ouzrout, "Evolution of 3C cyber-physical systems architecture for industry 4.0," in *Service Orientation in Holonic and Multi-Agent Manufacturing Cham*, T. Borangiu, D. Trentesaux, A. Thomas, and S. Cavalieri, Eds., pp. 448–459, Springer International Publishing, Cham, Switzerland, 2019.

[32] A. J. H. Redelinghuys, A. H. Basson, and K. Kruger, "A six-layer architecture for the digital twin: a manufacturing case study implementation," *Journal of Intelligent Manufacturing*, vol. 31, no. 6, pp. 1383–1402, 2020.

[33] E. Kavakli, J. Buenabad-Chávez, V. Tountopoulos, P. Loucopoulos, and R. Sakellariou, "Specification of a software architecture for an industry 4.0 environment," in *Proceedings of the 2018 6th International Conference on Enterprise Systems ES*, pp. 36–43, Limassol, Cyprus, October 2018.

[34] A. K. Sahu, A. K. Sahu, and N. K. Sahu, "A review on the research growth of industry 4.0," *International Journal of Business Analytics*, vol. 7, no. 1, pp. 77–97, 2020.

[35] O. Nwauka, A. Telukdarie, and J. Enslin, "Virtual power plant basic requirements for integration of distributed energy resources, driven by industry 4.0," in *Proceedings of the International Conference on Industrial Engineering and Operations Management; 2018*, pp. 511–523, Paris, France, July, 2018.

[36] M. Ringhofer, G. Wimmer, J. F. Plaul, E. Tatschl-Unterberger, and K. Herzog, "Transition of the steelmaking industry to digital technologies," *Chernye Metally*, no. 3, pp. 12–17, 2018.

[37] I. Hardai, B. Illés, and Á. Bányai, "Efficiency improvement of reverse logistics in industry 4.0 environment," in *Proceedings of the 1st International Conference on Engineering Solutions for Sustainable Development (ICESSD 2019)*, pp. 169–177, Miskolc, Hungary, October 2019.

[38] D. Ivanov, A. Dolgui, A. Das, and B. Sokolov, "Digital supply chain twins: managing the ripple effect, resilience, and disruption risks by data-driven optimization, simulation, and visibility," in *Handbook of Ripple Effects in the Supply Chain*, D. Ivanov, A. Dolgui, and B. Sokolov, Eds., Springer International Publishing, Cham, Switzerland, pp. 309–332, 2019.

[39] O. Protalinsky, A. Khanova, and I. Shcherbatov, "Simulation of power assets management process," in *Recent Research in Control Engineering and Decision Making Cham*, O. Dolinina, A. Brovko, V. Pechenkin, A. Lvov, V. Zhmud,

and V. Kreinovich, Eds., Springer International Publishing, Cham, Switzerland, pp. 488–501, 2019.

[40] E. Hozdić, "Smart factory for industry 4.0: a review," *International Journal of Modern Manufacturing Technologies*, vol. 7, no. 1, pp. 28–35, 2015.

[41] D. Sun, R. Huang, Y. Chen et al., "PlanningVis: a visual analytics approach to production planning in smart factories," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, pp. 579–589, 2020.

[42] J. Wan, M. Yi, D. Li, C. Zhang, S. Wang, and K. Zhou, "Mobile services for customization manufacturing systems: an example of industry 4.0," *IEEE Access*, vol. 4, pp. 8977–8986, 2016.

[43] R. Rolle, V. Martucci, and E. Godoy, "Architecture for digital twin implementation focusing on industry 4.0," *IEEE Latin America Transactions*, vol. 18, no. 5, pp. 889–898, 2020.

[44] T. Nishi, M. Matsuda, M. Hasegawa, R. Alizadeh, Z. Liu, and T. Terunuma, "Automatic construction of virtual supply chain as multi-agent system using enterprise E-catalogues," *International Journal of Automation Technology*, vol. 14, no. 5, pp. 713–722, 2020.

[45] X. Ma, J. Wang, Q. Bai, and S. Wang, "Optimization of a three-echelon cold chain considering freshness-keeping efforts under cap-and-trade regulation in Industry 4.0," *International Journal of Production Economics*, vol. 220, p. 107457, 2020.

[46] E. Chacon, J. Albarrán, and L. Cruz Salazar, "The control of water distribution systems as a holonic system," in *Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future*, pp. 352–365, Springer International Publishing, Cham, Switzerland, 2020.

[47] L. Pérez, S. Rodríguez-Jiménez, N. Rodríguez, R. Usamentiaga, and D. F. García, "Digital twin and virtual reality based methodology for multi-robot manufacturing cell commissioning," *Applied Sciences*, vol. 10, no. 10, p. 3633, 2020.

[48] V. K. Chawla, S. Angra, S. Suri, and R. S. Kalra, "A synergic framework for cyber-physical production systems in the context of industry 4.0 and beyond," *International Journal of Data and Network Science*, vol. 4, pp. 237–244, 2020.

[49] M. Trstenjak, T. Opetuk, H. Cajner, and N. Tosanovic, "Process planning in industry 4.0-current state, potential and management of transformation," *Sustainability*, vol. 12, no. 15, p. 5878, 2020.

[50] N. I. Giannoccaro, G. Persico, S. Strazzella, A. Lay-Ekuakille, and P. Visconti, "A system for optimizing fertilizer dosing in innovative smart fertigation pipelines: modeling, construction, testing and control," *International Journal of Precision Engineering and Manufacturing*, vol. 21, no. 1, pp. 1581–1596, 2020.

[51] M. Pekarcikova, P. Trebuna, M. Kliment, and L. Rosocha, "Material flow optimization through E-kanban system simulation," *International Journal of Simulation Modelling*, vol. 19, no. 2, pp. 243–254, 2020.

[52] D. Ivanov, B. Sokolov, F. Werner, and A. Dolgui, "Proactive scheduling and reactive real-time control in industry 4.0," in *Scheduling in Industry 4.0 and Cloud Manufacturing*, B. Sokolov, D. Ivanov, and A. Dolgui, Eds., pp. 11–37, Springer International Publishing, Cham, Switzerland, 2020.

[53] J. Aston and R. P. Freire, "The car as a transformer," in *Intelligent Human Systems Integration 2020*, T. Ahram, W. Karwowski, A. Vergnano, F. Leali, and R. Taiar, Eds., pp. 118–123, Springer International Publishing, Cham, Switzerland, 2020.

[54] V. Colla, V. Iannino, J. Denker, and M. Göttsche, "A CPS-based simulation platform for long production factories," *Metals-Open Access Metallurgy Journal*, vol. 9, no. 10, p. 1025, 2019.

[55] V. Slaný, A. Lučanský, P. Koudelka, J. Mareček, E. Krčálová, and R. Martínek, "An integrated IoT architecture for smart metering using next generation sensor for water management based on LoRaWAN technology: a pilot study," *Sensors*, vol. 20, no. 17, p. 4712, 2020.

[56] N. Gramegna, F. Greggio, and F. Bonollo, "Smart factory competitiveness based on real time monitoring and quality predictive model applied to multi-stages production lines," in *Advances in Production Management Systems. Towards Smart and Digital Manufacturing*, B. Lalic, V. Majstorovic, U. Marjanovic, G. von Cieminski, and D. Romero, Eds., pp. 185–196, Springer International Publishing, Cham, Switzerland, 2020.

[57] M. Sanchez, E. Exposito, and J. Aguilar, "Autonomic computing in manufacturing process coordination in industry 4.0 context," *Journal of Industrial Information Integration*, vol. 19, p. 100159, 2020.

[58] W. Meng, D. Hongyan, Z. Shiyuan, D. Zhankui, and W. Zige, "The application of the positive semi-definite kernel space for SVM in quality prediction," *Recent Advances in Computer Science and Communications*, vol. 13, no. 2, pp. 228–233, 2020.

[59] R. H. Fuad, J. M. Nilakantan, and N. Peter, "An integrated approach for line balancing and AGV scheduling towards smart assembly systems," *Assembly Automation*, vol. 40, no. 2, pp. 219–234, 2020.

[60] S. Khan, M. Farnsworth, R. McWilliam, and J. Erkoyuncu, "On the requirements of digital twin-driven autonomous maintenance," *Annual Reviews in Control*, vol. 50, pp. 13–28, 2020.

[61] T. Le Grand and R. Deneckere, "COOC: an agile change management method," in *Proceedings of the 2019 IEEE 21st Conference on Business Informatics (CBI)*, pp. 28–37, Moscow, Russia, July 2019.

[62] M. Ebrahimi, A. Baboli, and E. Rother, "A roadmap for evolution of existing production system toward the factory of the future: a case study in automotive industry," in *Proceedings of the 2018 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*, pp. 274–281, Marrakech, Morocco, November 2018.

[63] L. Mazzola, P. Waibel, P. Kaphanke, and M. Klusch, "Smart process optimization and adaptive execution with semantic services in cloud manufacturing," *Information*, vol. 9, no. 11, pp. 279–307, 2018.

[64] A. Legarretaetxebarria, M. Quartulli, I. Olaizola, and M. Serrano, "Optimal scheduling of manufacturing processes across multiple production lines by polynomial optimization and bagged bounded binary knapsack," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 11, no. 1, pp. 83–91, 2017.

[65] L. Banjanovic-Mehmedovic and A. Baluković, "PSO optimized fuzzy controller for mobile robot path tracking," in *New Technologies, Development and Application III Cham*, I. Karabegović, Ed., pp. 413–421, Springer International Publishing, Cham, Switzerland, 2020.

[66] R. Carli, S. Digiesi, M. Dotoli, and F. Facchini, "A control strategy for smart energy charging of warehouse material handling equipment," *Procedia Manufacturing*, vol. 42, pp. 503–510, 2020.

[67] M. Kaltenbrunner, M. A. Huka, and M. Gronalt, "Automating production planning and control in pallet manufacturing-a case study," *Procedia Manufacturing*, vol. 42, pp. 119–124, 2020.

[68] M. R. Islam, J. Al Rafi, and M. S. Hossain, "Operation planning of renewable energy based hybrid system incorporating waste-to-energy (WtE) technologies," in *Proceedings of the 2019 International Conference on Sustainable Technologies for Industry 4.0, STI*, pp. 1–5, Dhaka, Bangladesh, December 2019.

[69] M. Liu and X. Liu, "Machine scheduling with stochastic release and processing times," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 2116–2121, 2019.

[70] H. E. van Staden and R. N. Boute, "The effect of multi-sensor data on condition-based maintenance policies," *European Journal of Operational Research*, vol. 290, no. 2, pp. 585–600, 2020.

[71] Z. Ma, J. Korsgaard, and B. N. Jorgensen, "Optimization of greenhouse production process: an investigation of energy efficiency potentials," in *Proceedings of the 2019 6th International Conference on Dependable Systems and their Applications, DSA*, pp. 365–370, Harbin, China, January 2020.

[72] M. Nouiri, D. Trentesaux, and A. Bekrar, "Towards energy efficient scheduling of manufacturing systems through collaboration between cyber physical production and energy systems," *Energies*, vol. 12, no. 23, p. 4448, 2019.

[73] S. Mantovani, I. L. Presti, L. Cavazzoni, and A. Baldini, "Influence of manufacturing constraints on the topology optimization of an automotive dashboard," *Procedia Manufacturing*, vol. 11, pp. 1700–1708, 2017.

[74] V. Cardini, M. Eugeni, A. Boschetto et al., "An additive manufacturing redesign of a military aircraft equipment within a logistic 4.0 framework," in *Proceedings of the International Astronautical Congress, IAC*, Washington, DC, USA, October 2019.

[75] M. Leite, T. Pinto, and C. Alves, "A real-time optimization algorithm for the integrated planning and scheduling problem towards the context of industry 4.0," *FME Transactions*, vol. 47, no. 4, pp. 775–781, 2019.

[76] T. H. Hsu, L. C. Wang, and P. C. Chu, "Development of a cloud-based advanced planning and scheduling system," *Procedia Manufacturing*, vol. 17, pp. 427–434, 2018.

[77] T. Drews, P. Molenda, O. Oechsle, and J. Koller, "Manufacturing system optimization with lean methods, manufacturing process objectives and fuzzy logic controller design," *Procedia CIRP*, vol. 93, pp. 658–663, 2020.

[78] R. Ribeiro, C. Analide, and O. Belo, "Improving productive processes using a process mining approach," in *Trends and Advances in Information Systems and Technologies*, Á. Rocha, H. Adeli, L. P. Reis, and S. Costanzo, Eds., pp. 736–745, Springer International Publishing, Cham, Switzerland, 2018.

[79] K. Zhang, W. Li, Y. Han, Z. Geng, and C. Chu, "Production capacity identification and analysis using novel multivariate nonlinear regression: application to resource optimization of industrial processes," *Journal of Cleaner Production*, vol. 282, p. 124469, 2020.

[80] B. A. McCarl, H. Moskowitz, and H. Furtan, "Quadratic programming applications," *Omega*, vol. 5, no. 1, pp. 43–55, 1977.

[81] H. Liu, Y. Li, Z. Duan, and C. Chen, "A review on multi-objective optimization framework in wind energy forecasting techniques and applications," *Energy Conversion and Management*, vol. 224, p. 113324, 2020.

[82] N. V. Sahinidis, "Optimization under uncertainty: state-of-the-art and opportunities," *Computers & Chemical Engineering*, vol. 28, no. 6-7, pp. 971–983, 2004.

[83] C. A. Petri, "Kommunikation mit automaten," Ph. D. Thesis, Institut für Instrumentelle Mathematik, Bonn, Germany, 1962.

[84] F. Friedler, K. Tarján, Y. W. Huang, and L. T. Fan, "Graph-theoretic approach to process synthesis: axioms and theorems," *Chemical Engineering Science*, vol. 47, no. 8, pp. 1973–1988, 1992.

[85] E. Kondili, C. C. Pantelides, and R. W. H. Sargent, "A general algorithm for short-term scheduling of batch operations-I. MILP formulation," *Computers & Chemical Engineering*, vol. 17, no. 2, pp. 211–227, 1993.

[86] M. Liu and X. Liu, "Profit-driven stochastic scheduling considering discounted cash flows under industry 4.0," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 2122–2127, 2019.

[87] R. B. Kuriakose and H. J. Vermaak, "Optimization of a real time web enabled mixed model stochastic assembly line to reduce production time," *Communications in Computer and Information Science*, vol. 1075, pp. 39–50, 2019.

[88] C. Müller, M. Grunewald, and T. S. Spengler, "Redundant configuration of robotic assembly lines with stochastic failures," *International Journal of Production Research*, vol. 56, no. 10, pp. 3662–3682, 2018.

[89] P. Marangé, D. Lemoine, A. Aubry et al., "Coupling robust optimization and model-checking techniques for robust scheduling in the context of industry 4.0," in *Scheduling in Industry 4.0 and Cloud Manufacturing*, B. Sokolov, D. Ivanov, and A. Dolgui, Eds., pp. 103–124, Springer International Publishing, Cham, Switzerland, 2020.

[90] Q. Zaman and A. König, "Self-x integrated sensor circuits immune to measurement noise in the presence of input perturbation by using robust optimization," *Technisches Messen*, vol. 86, no. 1, pp. S107–S111, 2019.

[91] C. Hu, X. Liu, J. Lu, and C.-H. Wang, "Distributionally robust optimization for power trading of waste-to-energy plants under uncertainty," *Applied Energy*, vol. 276, p. 115509, 2020.

[92] J. Pierezan, G. Maidl, E. Massashi Yamao, L. dos Santos Coelho, and V. Cocco Mariani, "Cultural coyote optimization algorithm applied to a heavy duty gas turbine operation," *Energy Conversion and Management*, vol. 199, p. 111932, 2019.

[93] J. Del Ser, E. Osaba, D. Molina et al., "Bio-inspired computation: where we stand and what's next," *Swarm and Evolutionary Computation*, vol. 48, pp. 220–250, 2019.

[94] J. L. S. De Silva, D. B. De Mesquita, H. S. Moreira et al., "Industry 4.0 approach aligned with the challenges of grid-connected photovoltaic systems," in *Proceedings of the 2019 IEEE PES Conference on Innovative Smart Grid Technologies, ISGT Latin America*, Gramado, Brazil, September 2019.

[95] C. Kim and H. Oh, "Simulation-based sensor network optimal design for detecting fracture in pipeline systems," in *Proceedings of SPIE-the International Society for Optical Engineering, Volume 10973*, Denver, CO, USA, March 2019.

[96] M. Liu, Z. Liu, X. Liu, and F. Chu, "Entropy-based bi-objective disassembly line balancing problem," in *Proceedings of the 2019 International Conference on Industrial Engineering and Systems Management, IESM*, Shanghai, China, September 2019.

[97] M. P. Marcos, J. L. Pitarch, and C. de Prada, "Decision support system for a heat-recovery section with equipment

degradation," *Decision Support Systems*, vol. 137, p. 113380, 2020.

[98] D. Bechtsis, N. Tsolakis, M. Vouzas, and D. Vlachos, "Industry 4.0: sustainable material handling processes in industrial environments," *Computer Aided Chemical Engineering*, vol. 40, pp. 2281–2286, 2017.

[99] A. Pichler, S. Poledna, and S. Thurner, "Systemic risk-efficient asset allocations: minimization of systemic risk as a network optimization problem," *Journal of Financial Stability*, p. 100809, 2020.

[100] Y. Sun, N. V. Sahinidis, A. Sundaram, and M.-S. Cheon, "Derivative-free optimization for chemical product design," *Current Opinion in Chemical Engineering*, vol. 27, pp. 98–106, 2020.

[101] H. Gao, A. Waechter, I. A. Konstantinov, S. G. Arturo, and L. J. Broadbelt, "Application and comparison of derivative-free optimization algorithms to control and optimize free radical polymerization simulated using the kinetic Monte Carlo method," *Computers & Chemical Engineering*, vol. 108, pp. 268–275, 2018.

[102] I. Negrellos-Ortiz, A. Flores-Tlacuahuac, and M. A. Gutiérrez-Limón, "Dynamic optimization of a cryogenic air separation unit using a derivative-free optimization approach," *Computers & Chemical Engineering*, vol. 109, pp. 1–8, 2018.

[103] L. Benhamou, V. Giard, M. Khouloud, P. Fenies, and F. Fontane, "Reverse blending: an economically efficient approach to the challenge of fertilizer mass customization," *International Journal of Production Economics*, vol. 226, p. 107603, 2020.

[104] J. Hong, N. Van Duc Long, G. R. Harvianto, J. Haider, and M. Lee, "Design and optimization of multi-effect-evaporation-assisted distillation configuration for recovery of 2,3-butanediol from fermentation broth," *Chemical Engineering and Processing-Process Intensification*, vol. 136, pp. 107–115, 2019.

[105] Y. Jiang, E. Liese, S. E. Zitney, and D. Bhattacharyya, "Optimal design of microtube recuperators for an indirect supercritical carbon dioxide recompression closed Brayton cycle," *Applied Energy*, vol. 216, pp. 634–648, 2018.

[106] D. Erkin, E. Dugundji, G. Koole, J. Gromicho, and R. van der Mei, "Buffering locations in retail deliveries," *Procedia Computer Science*, vol. 151, pp. 116–123, 2019.

[107] S. M. Hassan, M. Azab, and A. Mokhtar, "Smart concrete transportation in semi-automated construction sites," in *Proceedings of the 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON*, pp. 661–667, Vancouver, Canada, October 2019.

[108] D. A. Chekired, L. Khoukhi, and H. T. Mouftah, "Industrial IoT data scheduling based on hierarchical fog computing: a key for enabling smart factory," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4590–4602, 2018.

[109] J. Jayakumar, J. Kandasamy, K. E. K. Vimal, and S. Hasibuan, "Modelling of sharing networks in the circular economy," *Journal of Modelling in Management*, vol. 15, no. 2, pp. 407–440, 2020.

[110] L. Bukata, P. Sucha, Z. Hanzalek, and P. Burget, "Energy optimization of robotic cells," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 92–102, 2017.

[111] Z. A. Çil, Z. Li, S. Mete, and E. Özceylan, "Mathematical model and bee algorithms for mixed-model assembly line balancing problem with physical human–robot collaboration," *Applied Soft Computing Journal*, vol. 93, p. 106394, 2020.

[112] B. Mihoubi, B. Bouzouia, and M. Gaham, "Reactive scheduling approach for solving a realistic flexible job shop scheduling problem," *International Journal of Production Research*, pp. 1–19, 2020.

[113] V. P. Zhalnin, A. S. Zakharova, D. A. Uzenkov, A. I. Vlasov, A. I. Krivoshein, and S. S. Filin, "Configuration-making algorithm for the smart machine controller based on the internet of things concept," *International Review of Electrical Engineering (IREE)*, vol. 14, no. 5, pp. 375–384, 2019.

[114] I. Maganha, C. Silva, N. Klement, A. B. dit Eynaud, L. Durville, and S. Moniz, "Hybrid optimisation approach for sequencing and assignment decision-making in reconfigurable assembly lines," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 1367–1372, 2019.

[115] P. Zhang, C. Wang, Z. Qin, and H. Cao, "A multidomain virtual network embedding algorithm based on multi-objective optimization for internet of drones architecture in industry 4.0," *Software-Practice and Experience*, pp. 1–19, 2020.

[116] G. H. A. Medeiros, Q. Cao, C. Zanni-Merk, and A. Samet, "Manufacturing as a service in industry 4.0: a multi-objective optimization approach," in *Intelligent Decision Technologies*, I. Czarnowski, R. J. Howlett, and L. C. Jain, Eds., Springer, Singapore, pp. 37–47, 2020.

[117] M. Afrin, J. Jin, A. Rahman, Y.-C. Tian, and A. Kulkarni, "Multi-objective resource allocation for edge cloud based robotic workflow in smart factory," *Future Generation Computer Systems*, vol. 97, pp. 119–130, 2019.

[118] P. Davidsson, J. A. Persson, and J. Holmgren, "On the integration of agent-based and mathematical optimization techniques," in *Agent and Multi-Agent Systems: Technologies and Applications*, N. T. Nguyen, A. Grzech, R. J. Howlett, and L. C. Jain, Eds., pp. 1–10, Springer, Berlin, Germany, 2007.

[119] H.-P. Wiendahl and P. Scholtissek, "Management and control of complexity in manufacturing," *CIRP Annals*, vol. 43, no. 2, pp. 533–540, 1994.

[120] C. C. Bozarth, D. P. Warsing, B. B. Flynn, and E. J. Flynn, "The impact of supply chain complexity on manufacturing plant performance," *Journal of Operations Management*, vol. 27, no. 1, pp. 78–93, 2009.

[121] A. V. Deshmukh, J. J. Talavage, and M. M. Barash, "Complexity in manufacturing systems, part 1: analysis of static complexity," *IIE Transactions*, vol. 30, no. 7, pp. 645–655, 1998.

[122] G. Ausiello, A. Marchetti-Spaccamela, P. Crescenzi, G. Gambosi, M. Protasi, and V. Kann, *Complexity and Approximation-Combinatorial Optimization Problems and their Approximability Properties*, Springer, Berlin, Germany, 1999.

[123] M. Ghaleb, H. Zolfagharinia, and S. Taghipour, "Real-time production scheduling in the industry-4.0 context: addressing uncertainties in job arrivals and machine breakdowns," *Computers & Operations Research*, vol. 123, p. 105031, 2020.

[124] H. ElMaraghy, A. Azab, G. Schuh, and C. Pulz, "Managing variations in products, processes and manufacturing systems," *CIRP Annals*, vol. 58, no. 1, pp. 441–446, 2009.

[125] Y. Koren, *The Global Manufacturing Revolution: Product-Process-Business Integration and Reconfigurable Systems*, John Wiley & Sons, Hoboken, NJ, USA, 2010.

[126] D. A. Rossit, F. Tohmé, and M. Frutos, "Production planning and scheduling in cyber-physical production systems: a

review," *International Journal of Computer Integrated Manufacturing*, vol. 32, no. 4-5, pp. 385–395, 2019.

[127] F. Sahin, A. Narayanan, and E. P. Robinson, "Rolling horizon planning in supply chains: review, implications and directions for future research," *International Journal of Production Research*, vol. 51, no. 18, pp. 5413–5436, 2013.

[128] C. Cimino, E. Negri, and L. Fumagalli, "Review of digital twin applications in manufacturing," *Computers in Industry*, vol. 113, p. 103130, 2019.

[129] J. Mertens, M. Challenger, K. Vanherpen, and J. Denil, "Towards real-time cyber-physical systems instrumentation for creating digital twins," in *Proceedings of the 2020 Spring Simulation Conference (SpringSim)*, pp. 1–12, IEEE, Fairfax, VA, USA, May 2020.

[130] G. Zhou, C. Zhang, Z. Li, K. Ding, and C. Wang, "Knowledge-driven digital twin manufacturing cell towards intelligent manufacturing," *International Journal of Production Research*, vol. 58, no. 4, pp. 1034–1051, 2020.

[131] F. Caputo, A. Greco, M. Fera, and R. Macchiaroli, "Digital twins to enhance the integration of ergonomics in the workplace design," *International Journal of Industrial Ergonomics*, vol. 71, pp. 20–31, 2019.

[132] T. Gabor, L. Belzner, M. Kiermeier, M. T. Beck, and A. Neitz, "A simulation-based architecture for smart cyber-physical systems," in *Proceedings of the 2016 IEEE International Conference on Autonomic Computing (ICAC)*, pp. 374–379, IEEE, Wurzburg, Germany, July 2016.

[133] J. De Tommaso, F. Rossi, N. Moradi, C. Pirola, G. S. Patience, and F. Galli, "Experimental methods in chemical engineering: process simulation," *The Canadian Journal of Chemical Engineering*, vol. 98, no. 11, pp. 2301–2320, 2020.

[134] Y. Chen, O. Yang, C. Sampat, P. Bhalode, R. Ramachandran, and M. Ierapetritou, "Digital twins in pharmaceutical and biopharmaceutical manufacturing: a literature review," *Processes*, vol. 8, no. 9, p. 1088, 2020.

[135] K. Židek, J. Piteľ, M. Adámek, P. Lazorík, and A. Hošovskỳ, "Digital twin of experimental smart manufacturing assembly system for industry 4.0 concept," *Sustainability*, vol. 12, no. 9, p. 3658, 2020.

[136] K. Y. H. Lim, P. Zheng, C.-H. Chen, and L. Huang, "A digital twin-enhanced system for engineering product family design and optimization," *Journal of Manufacturing Systems*, vol. 57, pp. 82–93, 2020.

[137] R. Batres, "Ontologies in process systems engineering," *Chemie Ingenieur Technik*, vol. 89, no. 11, pp. 1421–1431, 2017.

[138] J. Wan, B. Yin, D. Li, A. Celesti, F. Tao, and Q. Hua, "An ontology-based resource reconfiguration method for manufacturing cyber-physical systems," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 6, pp. 2537–2546, 2018.

[139] H. Tang, D. Li, S. Wang, and Z. Dong, "CASOA: an architecture for agent-based manufacturing system in the context of industry 4.0," *IEEE Access*, vol. 6, pp. 12746–12754, 2017.

[140] J. Pękala, "Data transformation using custom class generator as part of systems integration in manufacturing company," in *Proceedings of the International Conference on Computer Networks*, pp. 397–409, Brunów, Poland, June 2016.

[141] U. Kannengiesser, M. Neubauer, and R. Heininger, "Integrating business processes and manufacturing operations based on S-BPM and B2MML," in *Proceedings of the 8th International Conference on Subject-Oriented Business Process Management*, pp. 1–10, Erlangen, Germany, April 2016.

[142] I. Harjunkoski, M. Hollender, R. Bauer, J. Doppelhamer, and S. Subbiah, "An easy-to-use and flexible scheduling component," *ABB Review*, vol. 3, pp. 92–97, 2017.

[143] M. Fechter, C. Seeber, and S. Chen, "Integrated process planning and resource allocation for collaborative robot workplace design," *Procedia CIRP*, vol. 72, pp. 39–44, 2018.

[144] G. Thiele, N. Khorsandi, and J. Krüger, "Energy efficiency optimization using AutomationML modeling and an EnPI methodology," in *Proceedings of the 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 1218–1221, IEEE, Zaragoza, Spain, September 2019.

[145] M. Fechter and A. Neb, "From 3D product data to hybrid assembly workplace generation using the AutomationML exchange file format," *Procedia CIRP*, vol. 81, pp. 57–62, 2019.

[146] P. Novák, F. J. Ekaputra, and S. Biffl, "Generation of simulation models in MATLAB-simulink based on AutomationML plant description," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7613–7620, 2017.

[147] A. Mazak, A. Lüder, S. Wolny et al., "Model-based generation of run-time data collection systems exploiting AutomationML," *at-Automatisierungstechnik*, vol. 66, no. 10, pp. 819–833, 2018.

Hindawi
Complexity
Volume 2021, Article ID 6642795, 7 pages
https://doi.org/10.1155/2021/6642795

Research Article

Adaptive Sliding Mode Control for a Class of Manipulator Systems with Output Constraint

Guangshi Li

Computer Center, Anshan Normal University, Anshan 114016, China

Correspondence should be addressed to Guangshi Li; wslgs@netease.com

Received 19 December 2020; Revised 19 January 2021; Accepted 25 January 2021; Published 3 February 2021

Academic Editor: Qiuye Sun

In this paper, an adaptive sliding mode control method based on neural networks is presented for a class of manipulator systems. The main characteristic of the discussed system is that the output variable is required to keep within a constraint set. In order to ensure that the system output meets the time-varying constraint condition, the asymmetric barrier Lyapunov function is selected in the design process. According to Lyapunov stability theory, the stability of the closed-loop system is analyzed. It is demonstrated that all signals in the resulted system are bounded, the tracking error converges to a small compact set, and the system output limits in its constrained set. Finally, the simulation example is used to show the effectiveness of the presented control strategy.

1. Introduction

The application of manipulator is of great significance to the development of human society. So far, many effective control strategies have been produced for manipulator control, such as the PID control [1, 2], the robust control [3–5], and the sliding mode control [6–8]. Among them, the PID control is more suitable for the control theory of other technologies difficult to adopt, or the mathematical models are not more accurate. In [9], a new adaptive neural network control method, which is based on a simple structure-like PID control, is constructed to deal with the robot system with incomplete model and disturbance. The feature of the robust control is that once the controller is designed, its parameters are fixed; that is, it is not affected by external interference, and it can guarantee the control performance. In [10, 11], the robust control approaches are proposed to solve the position tracking problem of the manipulator systems. However, if the function of the considered system needs to be improved or made major changes, then the controller requires to be redesigned, which cannot be realized for the robust control method. The sliding mode control has the good stability against external disturbance and system unmodeled dynamics. For this, the sliding mode control will be considered in this paper.

In recent years, with the rapid development of science and technique, the adaptive control technology has become particularly important. Especially, the fuzzy logic system-based and neural network-based adaptive control is one of the current research hot topics. These two methods are more suitable for nonlinear systems with unknown dynamics. In this aspect, a large number of research results have been obtained [12–15]. Concrete results such as that in [16], a neural network-based adaptive FTC controller was proposed for the nonstrict-feedback interconnected large-scale systems, in which the neural networks were used to approximate the unknown terms of the considered systems. In [17], the fuzzy logic systems were presented to approximate the unknown terms, and then an adaptive fuzzy controller was designed by using backstepping technique, which ensured the finite-time convergence.

In addition, under the influence of environment and other factors, variables in the controlled system are often limited [18–20]; that is, variables need to be within a certain range and cannot be arbitrarily large. Therefore, constraint control has been widely concerned by researchers, and abundant achievements have been obtained [21–24]. In [25], the nonstrict feedback system with full-state constraints was considered, as well as input saturation and unmodeled dynamics. By introducing a logarithm nonlinear mapping,

the considered system with full-state constraints was transformed into an unconstrained one.

According to the above analysis, an adaptive sliding mode controller with time-varying output constraint is designed. In order to ensure that the system output is with the specified time-varying constraint range, the asymmetric logarithmic Lyapunov function is selected. The unknown nonlinear function existed in the considered system is approximated by the radial basis function neural networks (RNFNNs). Then, an adaptive sliding mode controller is constructed by combining the backstepping technique and the sliding mode control approach.

Compared with the existing results, the innovations of this paper mainly include the following two aspects:

(1) Compared with the existing results in [21–25], where the constrained problems were considered and the symmetric barrier Lyapunov functions were employed, in this paper, the asymmetric barrier Lyapunov functions were used to solve the output constraint problem. Compared with symmetric barrier Lyapunov functions, asymmetric barrier Lyapunov functions are more widely used and less conservative.

(2) A new adaptive sliding mode control approach for manipulator systems is proposed by combining the barrier Lyapunov function method and the backstepping technique. A sufficient condition for the adaptive constrained control problem is derived. Moreover, the uncertainties are considered in this paper, which increase the design complexity.

The rest of this paper is organized as follows: Section 2 presents the problem description which includes the manipulator system description, the required assumption, and the control objective. Section 3 shows the design process of the adaptive NN controller, and the main result is also proposed. Section 4 contains the simulation example, and Section 5 is conclusion.

2. Problem Description

Consider the manipulator system in [26]

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = \dfrac{1}{M}x_3 - \dfrac{N}{M}\sin x_1 - \dfrac{B}{M}x_2, \\ \dot{x}_3 = \dfrac{1}{L}u - \dfrac{K_B}{L}x_2 - \dfrac{R}{L}x_3, \\ y = x_1, \end{cases} \tag{1}$$

where x_1, x_2, x_3 are the system states, $u \in R$ and $y \in R$ are the control input and output, respectively, and the definitions of related notations can be found in [26]. In this paper, the system output is required to satisfy a certain constraint condition. That is to say the system output $y(t)$ satisfies the constraint condition that $\underline{k}(t) \le y(t) \le \overline{k}(t)$, where $\underline{k}(t)$ and $\overline{k}(t)$ are continuous functions.

Assumption 1 (see [27]). The desired reference signal $y_d(t)$ is bounded, which means that there exists a constant $\overline{Y}$ such that $|y_d(t)| \le Y$.

The control target of this paper is to design an adaptive NN controller such that the closed-loop system is stable, and the system output can track the desired reference signal $y_d(t)$ as well as satisfy its constraint condition.

3. The Adaptive NN Controller Design

In this paper, the backstepping method is used to design the adaptive NN controller, and the following coordinate transformation is defined:

$$\begin{cases} z_1 = x_1 - y_d, \\ z_2 = x_2 - \alpha_1, \\ z_3 = x_3 - \alpha_2, \end{cases} \tag{2}$$

where z_1, z_2, and z_3 are the error variables and α_1 and α_2 are the virtual controllers, which will be designed in the later procedure.

Step 1. From (1) and (2), one gets

$$\begin{aligned} \dot{z}_1 &= \dot{x}_1 - \dot{y}_d \\ &= z_2 + \alpha_1 - \dot{y}_d. \end{aligned} \tag{3}$$

Consider the following asymmetric obstacle Lyapunov function:

$$V_1 = \frac{q(z_1)}{2}\log\frac{k_{z_1}^2(t)}{k_{z_1}^2(t) - z_1^2} + \frac{1 - q(z_1)}{2}\log\frac{k_{z_2}^2(t)}{k_{z_2}^2(t) - z_1^2}, \tag{4}$$

where $q(z_1)$ is a piecewise function, which means that if $z_1 > 0$, then $q(z_1) = 1$, otherwise $q(z_1) = 0$. $k_{z_1}(t)$ and $k_{z_2}(t)$ are positive definite smooth bounded function, which satisfy $k_{z_1}(t) = \overline{k}(t) - Y$ and $k_{z_2}(t) = -\underline{k}(t) - Y$, respectively.

For convenience of expression, define the following error coordinate transformation:

$$\begin{cases} e_1 = \dfrac{z_1}{k_{z_1}}, \\ e_2 = \dfrac{z_1}{k_{z_2}}, \\ e = q(z_1)e_1 + (1 - q(z_1))e_2, \end{cases} \tag{5}$$

where e_1, e_2, and e are transfer variables. Based on the above transformation, we know that if $z_1 > 0$, then $q(z_1) = 1$ and $e = e_1$. Furthermore,

$$V_1 = \frac{1}{2}\log\frac{1}{1 - e_1^2} = \frac{1}{2}\log\frac{1}{1 - e^2}. \tag{6}$$

Similarly, if $z_1 \le 0$, then $q(z_1) = 0$ and $e = e_2$. Thus we have

$$V_1 = \frac{1}{2}\log\frac{1}{1-e_2^2} = \frac{1}{2}\log\frac{1}{1-e^2}. \tag{7}$$

Combining (6) and (7), one has

$$V_1 = \frac{1}{2}\log\frac{1}{1-e^2}. \tag{8}$$

Then, the time derivative of V_1 is obtained as

$$\dot{V}_1 = \frac{q(z_1)e_1}{k_{z_1}\left(1-e_1^2\right)}\left(z_2 + \alpha_1 - \dot{y}_d - z_1\frac{\dot{k}_{z_1}}{k_{z_1}}\right) + \frac{(1-q(z_1))e_2}{k_{z_2}\left(1-e_2^2\right)}\left(z_2 + \alpha_1 - \dot{y}_d - z_1\frac{\dot{k}_{z_2}}{k_{z_2}}\right). \tag{9}$$

Design the following virtual controller:

$$\alpha_1 = -\left(\gamma_1 + \overline{\gamma}_1(t)\right)z_1 + \dot{y}_d, \tag{10}$$

where $\gamma_1 > 0$ is a constant and $\overline{\gamma}_1(t) = \sqrt{(\dot{k}_{z_1}/k_{z_1})^2 + (\dot{k}_{z_2}/k_{z_2})^2 + \gamma_0}$ is a time-varying gain with γ_0 being a positive constant.

Substituting (10) into (8), the following holds:

$$\dot{V}_1 \le -\frac{\gamma_1 e^2}{1-e^2} + \left(\frac{q(z_1)}{k_{z_1}^2 - z_1^2} + \frac{1-q(z_1)}{k_{z_2}^2 - z_1^2}\right)z_1 z_2 \le -\frac{\gamma_1 e^2}{1-e^2} + \mu z_1 z_2, \tag{11}$$

where $\mu = (q(z_1)/k_{z_1}^2 - z_1^2) + (1 - q(z_1)/k_{z_2}^2 - z_1^2)$.

Step 2. From (1) and (2), $\dot{z}_2$ can be described as

$$\dot{z}_2 = \frac{1}{M}\left(x_3 - N\ \sin\ x_1 - Bx_2 - M\dot{\alpha}_1\right). \tag{12}$$

In this paper, we consider the case that the load mass is uncertain, and then according to the definitions of the system parameter in [26], it is easy to know that M and N are unknown, which implies that the function $-N\ \sin\ x_1 - Bx_2 - M\dot{\alpha}_1$ is also unknown. Here, the RBFNNs are used to approximate it, i.e.,

$$-N\ \sin\ x_1 - Bx_2 - M\dot{\alpha}_1 = \omega_2^{*T}\varphi_2(X_2) + \varepsilon_2(X_2), \tag{13}$$

where ω_2^{*T} stands for the ideal weight vector of the NNs, $\varphi_2(X_2)$ is a Gaussian function with $X_2 = [x_1, x_2, x_3, y_d, \dot{y}_d]^T$ being the NN input, and $\varepsilon(X_2)$ is the approximation error. In addition, $\|\varepsilon_2(X_2)\| \le \overline{\varepsilon}_2$.

Assumption 2 (see [28]). The ideal weight vectors of RBFNNs are bounded.

Substituting (13) into (12), one gets

$$\dot{z}_2 = \frac{1}{M}\left(x_3 + \omega_2^{*T}\varphi_2(X_2) + \varepsilon_2(X_2)\right). \tag{14}$$

Select the following Lyapunov function:

$$V_2 = V_1 + \frac{M}{2}z_2^2 + \frac{1}{2\delta_2}\widetilde{W}_2^2, \tag{15}$$

where $\delta_2 > 0$ is a design parameter. $\widetilde{W}_2 = W_2^* - \widehat{W}_2$ is the estimation error with $\widehat{W}_2$ being the estimation of W_2^*, and $W_2^* = \|\omega_2^*\|^2$.

Then, combining (11) and (14), the time derivative of V_2 is

$$\dot{V}_2 = \dot{V}_1 + Mz_2\dot{z}_2 - \frac{1}{\delta_2}\widetilde{\omega}_2^T\dot{\widehat{\omega}}_2 \le -\frac{\gamma_1 e^2}{1-e^2} + \mu z_1 z_2 + z_2\left(x_3 + \omega_2^{*T}\varphi_2(X_2) + \varepsilon_2(X_2)\right) - \frac{1}{\delta_2}\widetilde{W}_2\dot{\widehat{W}}_2. \tag{16}$$

Based on Young's inequality, it has

$$z_2\omega_2^{*T}\varphi_2(X_2) \le \frac{z_2^2 W_2^*\varphi_2^T(X_2)\varphi_2(X_2)}{4\xi_2} + \xi_2 z_2\varepsilon_2(X_2) \le \frac{1}{2}z_2^2 + \frac{1}{2}\overline{\varepsilon}_2^2, \tag{17}$$

where $\xi_2 > 0$ is a parameter. Then, the derivative of V_2 becomes

$$\dot{V}_2 \le -\frac{\gamma_1 e^2}{1-e^2} + z_2\left(z_3 + \alpha_2 + \mu z_1 + \frac{1}{2}z_2 + \frac{z_2\widehat{W}_2\varphi_2^T(X_2)\varphi_2(X_2)}{4\xi_2}\right) + \widetilde{W}_2\left(\frac{z_2^2\varphi_2^T(X_2)\varphi_2(X_2)}{4\xi_2} - \frac{1}{\delta_2}\dot{\widehat{W}}_2\right) + \frac{1}{2}\overline{\varepsilon}_2^2 + \xi_2. \tag{18}$$

Select the following virtual controller and adaptive law

$$\alpha_2 = -\gamma_2 z_2 - \mu z_1 - \frac{1}{2}z_2 - \frac{z_2\widehat{W}_2\varphi_2^T(X_2)\varphi_2(X_2)}{4\xi_2}, \tag{19}$$

$$\dot{\widehat{W}}_2 = \delta_2\left(\frac{z_2^2\varphi_2^T(X_2)\varphi_2(X_2)}{4\xi_2} - \sigma_2\widehat{W}_2\right), \tag{20}$$

where γ_2 and σ_2 are the positive constants.

Substituting (19) and (20) into $\dot{V}_2$, one has

$$\dot{V}_2 \le -\frac{\gamma_1 e^2}{1-e^2} - \gamma_2 z_2^2 + z_2 z_3 + \sigma_2\widetilde{W}_2\widehat{W}_2 + \frac{1}{2}\overline{\varepsilon}_2^2 \le -\frac{\gamma_1 e^2}{1-e^2} - \gamma_2 z_2^2 - \frac{1}{2}\sigma_2\widetilde{W}_2^2 + z_2 z_3 + \frac{1}{2}\sigma_2 W_2^{*2} + \frac{1}{2}\overline{\varepsilon}_2^2 + \xi_2. \tag{21}$$

Step 3. In order to ensure that the tracking error of the system tends to zero in finite time, improve the convergence

accuracy, and avoid singular problems; an integral end face sliding surface is designed as

$$s = z_3 + \eta \int_0^t z_3^{(l/n)} \mathrm{d}\omega, \tag{22}$$

where $\eta > 0$, $l > 0$, $n > 0$, and $1 < (l/n) < 2$.

According to $z_3 = x_3 - \alpha_2$ and $\dot{x}_3 = (1/L)u - (K_B/L)x_2 - (R/L)x_3$, $\dot{s}$ can be obtained as

$$\dot{s} = \frac{1}{L}\left(u - K_B x_2 - R x_3 - L\dot{\alpha}_2 + L\eta z_3^{(l/n)}\right). \tag{23}$$

The function $-K_B x_2 - R x_3 - L\dot{\alpha}_2 + L\eta z_3^{(l/n)}$ is unknown due to the unknown parameter K_B, and thus the RBFNNs are used to approximate it

$$-K_B x_2 - R x_3 - L\dot{\alpha}_2 + L\eta z_3^{(l/n)} = \omega_3^{*T}\varphi_3(X_3) + \varepsilon_3(X_3). \tag{24}$$

Therefore, $\dot{s}$ becomes

$$\dot{s} = \frac{1}{L}\left(u + \omega_3^{*T}\varphi_3(X_3) + \varepsilon_3(X_3)\right). \tag{25}$$

Consider the Lyapunov function

$$V_3 = V_2 + \frac{L}{2}s^2 + \frac{1}{2\delta_3}\tilde{W}_3^2, \tag{26}$$

where δ_3 is a positive design parameter. $\tilde{W}_3 = W_3^* - \hat{W}_3$ is the estimation error with $\hat{W}_3$ being the estimation of W_3^*, and $W_3^* = \|\omega_3^*\|^2$.

Taking the derivative of V_3 along time, one has

$$\dot{V}_3 = \dot{V}_2 + Ls\dot{s} - \frac{1}{\delta_3}\tilde{W}_3\dot{\hat{W}}_3. \tag{27}$$

Substituting (21) and (25) into (27), it has

$$\dot{V}_3 \le -\frac{\gamma_1 e^2}{1-e^2} - \gamma_2 z_2^2 - \frac{1}{2}\sigma_2\tilde{W}_2^2 + z_2 z_3 + \frac{1}{2}\sigma_2 W_2^{*2} + \frac{1}{2}\bar{\varepsilon}_2^2 + \xi_2 - \frac{1}{\delta_3}\tilde{W}_3\dot{\hat{W}}_3 + s\left(u + \omega_3^{*T}\varphi_3(X_3) + \varepsilon_3(X_3)\right). \tag{28}$$

Using Young's inequality, we have

$$s\omega_3^{*T}\varphi_3(X_3) \le \frac{s^2 W_3\varphi_3^T(X_3)\varphi_3(X_3)}{4\xi_3} + \xi_3 s\varepsilon_3(X_3) \le \frac{1}{2}s^2 + \frac{1}{2}\bar{\varepsilon}_3^2, \tag{29}$$

and then, we can deduce that

$$\dot{V}_3 \le -\frac{\gamma_1 e^2}{1-e^2} - \gamma_2 z_2^2 - \frac{1}{2}\sigma_2\tilde{W}_2^2 + \frac{1}{2}\sigma_2 W_2^{*2} + \frac{1}{2}\bar{\varepsilon}_2^2 + \frac{1}{2}\bar{\varepsilon}_3^2 + s\left(u + \frac{s\hat{W}_3\varphi_3^T(X_3)\varphi_3(X_3)}{4\xi_3} + \frac{1}{2}s + \frac{z_2 z_3}{s}\right) + \tilde{W}_3\left(\frac{s^2\varphi_3^T(X_3)\varphi_3(X_3)}{4\xi_3} - \frac{1}{\delta_3}\dot{\hat{W}}_3\right) + \xi_2 + \xi_3. \tag{30}$$

Design the following actual controller and adaptive law:

$$u = -\gamma_3 s - \frac{1}{2}s - \frac{z_2 z_3}{s} - \frac{s\hat{W}_3\varphi_3^T(X_3)\varphi_3(X_3)}{4\xi_3}, \tag{31}$$

$$\dot{\hat{W}}_3 = \delta_3\left(\frac{s^2\varphi_3^T(X_3)\varphi_3(X_3)}{4\xi_3} - \sigma_3\hat{W}_3\right), \tag{32}$$

where γ_3 and σ_3 are the positive parameters.

On the basis of (31) and (32), we get

$$\dot{V}_3 \le -\frac{\gamma_1 e^2}{1-e^2} - \gamma_2 z_2^2 - \gamma_3 s^2 - \frac{1}{2}\sum_{j=2}^{3}\sigma_j\tilde{W}_j^2 + \frac{1}{2}\sum_{j=2}^{3}\sigma_j W_j^{*2} + \frac{1}{2}\sum_{j=2}^{3}\left(\bar{\varepsilon}_j^2 + \xi_j\right). \tag{33}$$

Let $C = \min\{\gamma_1, \gamma_2/M, \gamma_3/L, \sigma_2\delta_2, \sigma_3\delta_3\}$, and $D = (1/2)\sum_{j=2}^{2}(\sigma_j\omega_j^{*T}\omega_j^* + \bar{\varepsilon}_j^2 + \xi_j)$, from which C and D are the constants.

Furthermore, consider the inequality $-(\gamma_1 e^2/1-e^2) \le -\gamma_1\log(1/1-e^2)$. Then, $\dot{V}_3$ can be described as

$$\dot{V}_3 \leq -CV_3 + D. \tag{34}$$

Based on the above analysis, the following theorem is obtained.

Theorem 1. *Consider the manipulator system (1); based on Assumptions 1-2, the controllers (10), (19), (31) and adaptive laws (20), (32) are designed such that all signals in the resulted system are bounded, the system output can track the reference signal to a compact set, and it does not violate its constraint.*

Proof. . Select the Lyapunov function as follows:

$$V = \frac{1}{2}\log\frac{1}{1-e^2} + \frac{M}{2}z_2^2 + \frac{L}{2}s^2 + \frac{1}{2}\sum_{j=2}^{3}\frac{1}{\delta_j}\tilde{W}_j^2. \tag{35}$$

According to the above analysis, it is easily known that

$$\dot{V} \leq -CV + D. \tag{36}$$

The above inequality is integrated on both sides over $[0, t]$, and then the following holds:

$$V(t) \leq V(0)e^{-Ct} + \frac{D}{C}. \tag{37}$$

According to the definition of V, we can know that e, z_2, z_3, $\tilde{\omega}_2$, and $\tilde{\omega}_3$ are all bounded for the bounded initial condition. When $z_1 > 0$, $e = e_1$, so e_1 is bounded. Since $e_1 = (z_1/k_{z_1})$ and k_{z_1} is a bounded function, z_1 is also bounded. Based on $z_1 = x_1 - y_d$ and the boundness of y_d, x_1 is bounded. According to the definition of α_1, it is easy to know α_1 is bounded. Furthermore, on the basis of $z_2 = x_2 - \alpha_1$, x_2 is also bounded. Similarly, α_2, x_3, and u are bounded. Thus, all signals in the resulted system are bounded.

In addition, from (35) and (36), one has

$$\frac{1}{2}\log\frac{1}{1-e^2} \leq V(0)e^{-Ct} + \frac{D}{C}. \tag{38}$$

Furthermore, when $z_1 > 0$, $e = e_1 = (z_1/k_{z_1})$, and then

$$z_1 \leq k_{z_1}\sqrt{1 - e^{-V(0)e^{-Ct} - (D/C)}}. \tag{39}$$

Similarly, when $z_1 \leq 0$, we have

$$z_1 \leq k_{z_2}\sqrt{1 - e^{-V(0)e^{-Ct} - (D/C)}}. \tag{40}$$

Therefore, when $t \longrightarrow \infty$, the tracking error z_1 converges to the compact set that

$$z_1 \leq k_z\sqrt{1 - e^{-(D/C)}}, \tag{41}$$

where $k_z = \min\{k_{z_1}, k_{z_2}\}$. That is to say the system output can track the reference signal to a compact set.

Furthermore, from $z_1 = x_1 - y_d$, one has $-|z_1| - |y_d| \leq |x_1| \leq |z_1| + |y_d|$. As stated at the beginning of this paper, we can deduce that when $z_1 > 0$, $x_1 \leq \bar{k}(t)$ and when $z_1 \leq 0$, $x_1 \geq \underline{k}(t)$. Thus, the system output satisfies the constraint condition. □

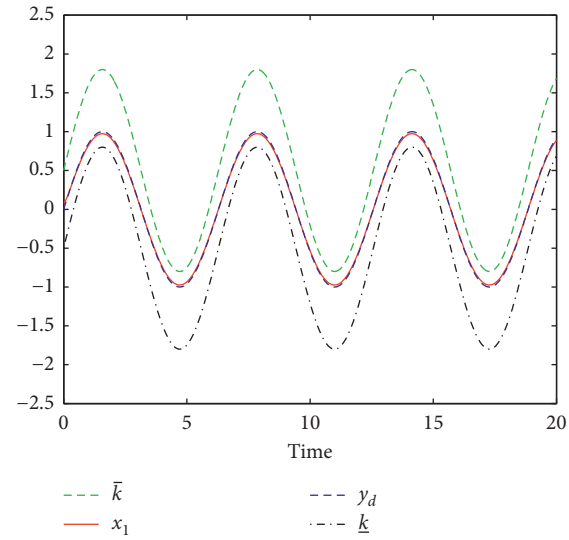


Figure 1: Trajectories of x_1, y_d and the constraint bounds of x_1.

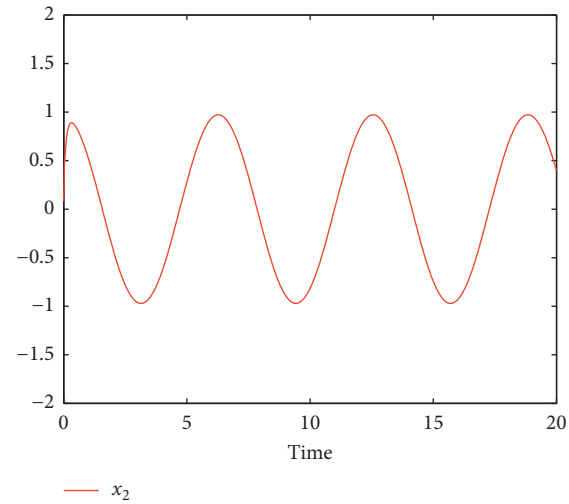


Figure 2: The trajectory of x_2.

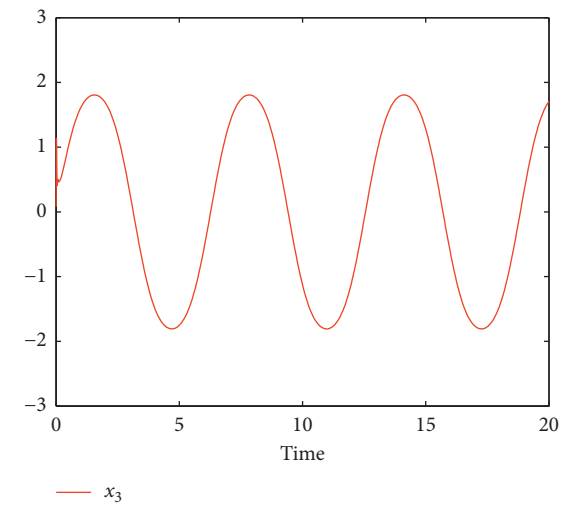


Figure 3: The trajectory of x_3.

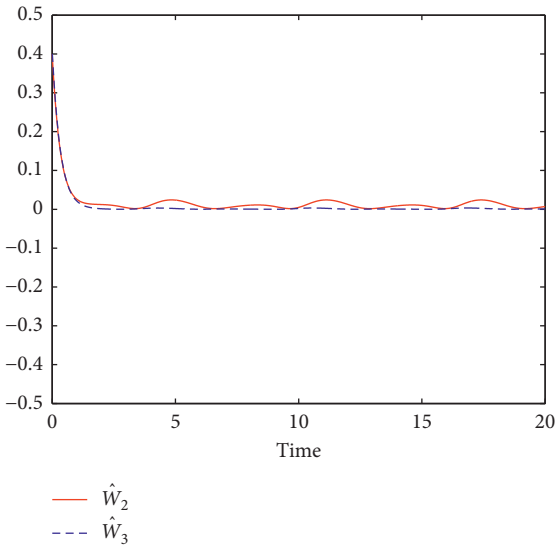


Figure 4: Trajectories of the update laws.

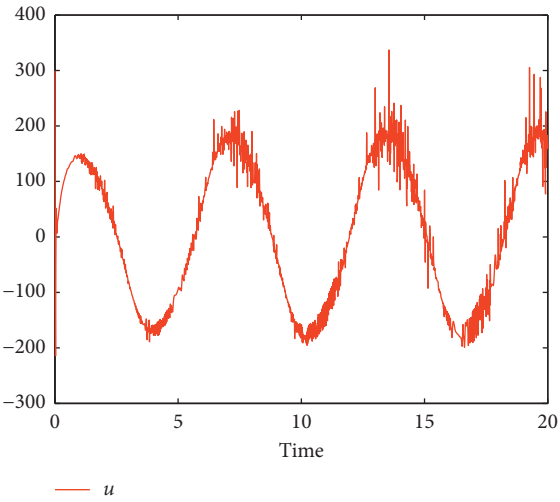


Figure 5: The trajectory of control input u.

This completes the proof of Theorem 1.

Remark 1. It is worth noting that equation (41) actually provides a direction on how to select some key design parameters: on the one hand, increasing γ_j, $j = 1, 2, 3$, σ_j, $j = 2, 3$, and δ_j, $j = 2, 3$ may help increase C, which further reduces (D/C). On the other hand, decreasing ξ_j, $j = 2, 3$, may help reduce D and reduces (D/C). The smaller (D/C) leads to larger $e^{-(D/C)}$, which implies that the tracking error z_1 can be smaller, and then the tracking performance is better.

4. Simulation Example

In this section, the manipulator system (1) is considered, where

$$M = \frac{J}{k_\tau} + \frac{mL_0^2}{3k_\tau} + \frac{M_0L_0^2}{k_\tau} + \frac{2M_0R_0^2}{5k_\tau},$$
$$B = \frac{B_0}{k_\tau}, \tag{42}$$
$$N = \frac{mL_0G}{2k_\tau} + \frac{M_0L_0G}{k_\tau},$$

with $J = 0.001625\text{kg} \cdot \text{m}^2$, $k_\tau = 0.9\,\text{N} \cdot \text{m/A}$, $m = 0.506\,\text{kg}$, $L_0 = 0.305\,\text{m}$, $M_0 = 0.434\,\text{kg}$, $R_0 = 0.022\,\text{m}$, $B_0 = 0.01625\,\text{N} \cdot \text{m} \cdot \text{s/rad}$, and $G = 9.8$, $L = 0.225\text{H}$.

In this simulation, the desired signal is chosen as $\underline{y}_d(t) = \sin t$, the constraint functions are $\bar{k}(t) = 1.3\sin(t) + 0.5$ and $\underline{k}(t) = 1.3\sin(t) - 0.5$. The initial values are selected as $x_1(0) = 0.001$, $x_2(0) = 0.012$, $x_3(0) = 0.001$, $W_2(0) = 0.01$, and $W_3(0) = 1$, and the other related design parameters are chosen as $\gamma_1 = 8$, $\gamma_2 = 8$, $\gamma_3 = 22$, $\gamma_0 = 0.3$, $\delta_2 = \delta_3 = 4$, $\sigma_2 = \sigma_3 = 3$, and $\xi_2 = \xi_3 = 1$.

Figures 1–5 show the simulation results of the adaptive sliding mode control strategy designed in this paper. Figure 1 describes the curves of system state x_1 (i.e., $y(t)$), the reference signal $y_d(t)$, and the constraint bounds of x_1. It can be seen that the system has a good tracking performance. At the same time, the output of the system is within the constraint range, i.e., it satisfies the constraint condition. Figures 2 and 3 show the trajectories of states x_2 and x_3, respectively. The adaptive laws for $\widehat{W}_2$ and $\widehat{W}_3$ are given in Figure 4. According to Figures 2–4, it is easily known that they are bounded. Figure 5 expresses the trajectory of the control input u, which is also bounded. Thus, the effectiveness of the proposed method is well explained.

5. Conclusion

A sliding mode-based adaptive neural network control strategy is proposed for a manipulator system with output constraint. Combining the sliding mode control method with backstepping technique, an adaptive controller is designed, which not only solves the output constraint issue of the considered system, but also makes the error between system output and the reference signal to converge to a small compact set. Moreover, the output satisfies the constraint condition. Finally, the effectiveness of the proposed approach is illustrated by the simulation example.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] V. M. Hernández-Guzmán, V. Santibáñez, and R. Campa, "PID control of robot manipulators equipped with brushless DC motors," *Robotica*, vol. 27, no. 2, pp. 225–233, 2009.

[2] V. M. Hernández-Guzmán and J. Orrante-Sakanassi, "Global PID control of robot manipulators equipped with PMSMs," *Asian Journal of Control*, vol. 20, no. 1, pp. 236–249, 2018.

[3] L. Liu and X. Li, "Event-triggered tracking control for active seat suspension systems with time-varying full state constraints," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–9, 2020.

[4] L. Tang and D. Li, "Time-varying barrier Lyapunov function based adaptive neural controller design for nonlinear pure-feedback systems with unknown hysteresis," *International Journal of Control, Automation and Systems*, vol. 17, no. 7, pp. 1642–1654, 2019.

[5] L. Tang, A. Chen, and D. Li, "Time-varying tan-type barrier Lyapunov function-based adaptive fuzzy control for switched systems with unknown dead zone," *IEEE Access*, vol. 7, no. 1, pp. 110928–110935, 2019.

[6] J. Wang, W. Chen, Z. Chen et al., "Neural terminal sliding-mode control for uncertain systems with building structure vibration," *Complexity*, vol. 2019, Article ID 507051, 9 pages, 2019.

[7] M. Van, "Adaptive neural integral sliding-mode control for tracking control of fully actuated uncertain surface vessels," *International Journal of Robust and Nonlinear Control*, vol. 29, no. 5, pp. 1537–1557, 2019.

[8] C. Huang, F. Naghdy, and H. Du, "Sliding mode predictive tracking control for uncertain steer-by-wire system," *Control Engineering Practice*, vol. 85, pp. 194–205, 2019.

[9] H. R. Nohooji, "Constrained neural adaptive PID control for robust manipulators," *Journal of the Franklin Institute*, vol. 357, no. 7, pp. 3907–3923, 2020.

[10] Y. Wang, X. Lai, P. Zhang, and M. Wu, "Adaptive robust control for planar n-link underactuated manipulator based on radial basis function neural network and online iterative correction method," *Journal of the Franklin Institute*, vol. 355, no. 17, pp. 8373–8391, 2018.

[11] S. Mishra, P. S. Londhe, S. Mohan, S. K. Vishvakarma, and B. M. Patre, "Robust task-space motion control of a mobile manipulator using a nonlinear control with an uncertainty estimator," *Computers & Electrical Engineering*, vol. 67, pp. 729–740, 2018.

[12] L. Liu, Y.-J. Liu, and S. Tong, "Fuzzy-based multierror constraint control for switched nonlinear systems and its applications," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 8, pp. 1519–1531, 2019.

[13] L. Liu, X. Li, Y. J. Liu, and S. Tong, "Neural network based adaptive event trigger control for a class of electromagnetic suspension systems," *Control Engineering Practice*, vol. 106, Article ID 104675, 2021.

[14] Y. Liu, X. Liu, Y. Jing, and Z. Zhang, "A novel finite-time adaptive fuzzy tracking control scheme for nonstrict feedback systems," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 4, pp. 646–658, 2019.

[15] L. Tang and J. Zhao, "Switched threshold-based fault detection for switched nonlinear systems with its application to Chuas circuit system," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 2, pp. 733–741, 2018.

[16] Y. Li and S. Tong, "Adaptive neural networks decentralized FTC design for nonstrict-feedback nonlinear interconnected large-scale systems against actuator faults," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 11, pp. 2541–2554, 2017.

[17] C.-W. Chang, C.-F. Hsu, and T.-T. Lee, "Backstepping-based finite-time adaptive fuzzy control of unknown nonlinear systems," *International Journal of Fuzzy Systems*, vol. 20, no. 8, pp. 2545–2555, 2018.

[18] W. Hu, C. Ruan, H. Nian, and D. Sun, "Zero-sequence current suppression strategy with common mode voltage control for open-end winding PMSM drives with common DC bus," *IEEE Transactions on Industrial Electronics*, p. 1, 2020.

[19] L. Liu, Y. J. Liu, S. C. Tong, and C. L. P. Chen, "Integral barrier Lyapunov function based adaptive control for switched nonlinear systems," *Science China Information Sciences*, vol. 63, no. 3, p. 132203, 2020.

[20] R. Wang, Q. Sun, P. Zhang et al., "Reduced-order transfer function model of the droop-controlled inverter via Jordan continued-fraction expansion," *IEEE Transactions on Energy Conversion*, vol. 35, no. 3, pp. 1585–1595, 2020.

[21] L. Wang and C. L. P. Chen, "Adaptive fuzzy dynamic surface control of nonlinear constrained systems with unknown virtual control coefficients," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 8, pp. 1737–1747, 2020.

[22] K. P. Tee and S. S. Ge, "Control of nonlinear systems with partial state constraints using a barrier Lyapunov function," *International Journal of Control*, vol. 84, no. 12, pp. 2008–2023, 2011.

[23] T. Gao, Y.-J. Liu, D. Li, S. Tong, and T. Li, "Adaptive neural control using tangent time-varying BLFs for a class of uncertain stochastic nonlinear systems with full state constraints," *IEEE Transactions on Cybernetics*, pp. 1–11, 2019.

[24] L. Wang, C. L. P. Chen, and H. Li, "Event-triggered adaptive control of saturated nonlinear systems with time-varying partial state constraints," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1485–1497, 2020.

[25] D. Ye, Y. Cai, H. Yang, and X. Zhao, "Adaptive neural-based control for non-strict feedback systems with full-state constraints and unmodeled dynamics," *Nonlinear Dynamics*, vol. 97, no. 1, pp. 715–732, 2019.

[26] D. M. Dawson, J. J. Carroll, and M. Schneider, "Integrator backstepping control of a brush DC motor turning a robotic load," *IEEE Transactions on Control Systems Technology*, vol. 2, no. 3, pp. 233–244, 1994.

[27] W. He, Y. Dong, and C. Sun, "Adaptive neural network control of unknown nonlinear affine systems with input deadzone and output constraint," *ISA Transactions*, vol. 58, pp. 96–104, 2015.

[28] D. J. Li, "Adaptive neural network control for a class of continuous stirred tank reactor systems," *Science China Information Sciences*, vol. 57, pp. 1–8, 2014.

Hindawi
Complexity
Volume 2021, Article ID 6662232, 14 pages
https://doi.org/10.1155/2021/6662232

Research Article

Vulnerability of Submarine Cable Network of Mainland China: Comparison of Vulnerability between before and after Construction of Trans-Arctic Cable System

Yongshun Xie[1,2] **and Chengjin Wang**[1,2]

[1]*Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China*
[2]*College of Resources and Environment, University of the Chinese Academy of Sciences, Beijing 100049, China*

Correspondence should be addressed to Chengjin Wang; cjwang@igsnrr.ac.cn

Received 3 November 2020; Revised 2 December 2020; Accepted 7 January 2021; Published 18 January 2021

Academic Editor: Qiuye Sun

The submarine optical-fiber cable (submarine cable) is a large connectivity infrastructure, which plays an important role in international communication, cyber-physical systems, and even national security. Although submarine cable network interruption may cause serious consequences, researching its vulnerability has not attracted much attention. This paper proposes a quantitative method to measure the vulnerability of the submarine cable network and evaluates the influence of the upcoming trans-Arctic cable (TAC) system on the submarine cable network of mainland China. To address this issue, first, the submarine cable network of mainland China is constructed. Further, methodology based on complex network and geospatial perspective is proposed to simulate the changes of network eigenvalues under different attacks and to quantitatively analyze the variation degree of network vulnerability. With the proposed method, the vulnerability of the submarine cable network of mainland China before and after construction of the trans-Arctic cable system is compared. The results reveal the key node countries and sea channels of the submarine cable network of mainland China and show the significance of the TAC system.

1. Introduction

As an important carrier and channel for information transmission between different regions, the submarine optical cable networks comprise large-scale connectivity infrastructures. However, the vulnerability of the submarine cable network is often ignored. Although plenty of research studies have paid attention to the topology structure and optimal deployment of the submarine cable network, there are few researches on its vulnerability. To fill this gap, this paper draws on the complex network theory and method to analyze the vulnerability of the submarine cable network of mainland China and compares the vulnerability between before and after construction of the trans-Arctic cable system.

The global submarine cable network, consisting of approximately 450 submarine cable systems, covering 1.2 million kilometers in length [1], and handling 99 percent of international data traffic, is the backbone of the physical infrastructure of the global Internet, as well as the most important information transmission medium in the world, currently and for the foreseeable future [2]. The coming decades may see the large-scale deployment of networked cyber-physical systems (CPSs), which puts forward new requirements for the stability and invulnerability of the submarine cable network. For the submarine cable network, it is not only the underlying physical system of CPSs but also the information flow it brings constitutes the cyber system of CPSs. Therefore, the vulnerability of submarine cable network is very important for CPSs and even national development in the era of Industry 4.0.

Vulnerability is an important concept, referring to the extent to which a system is affected when a disaster occurs; it is widely used in the study of ecology [3–5], disaster management [6, 7], and urban economics [8, 9]. Watts and Strogatz [10] and Barabasi and Albert [11] found the small

world characteristics and scale-free characteristics of the real world and opened a new era of complex network system research. Subsequently, in the early 2000s, Albert et al. [12], Broder et al. [13], and Bollobás and Riordan [14] explored the vulnerability of complex networks with different characteristics; this work laid the foundation for later scholars to explore the vulnerability of complex networks and became important in recognizing the key nodes and edges in complex networks.

In fact, many telecommunication networks have been found to have the characteristics of complex networks. For example, Schintler et al. [15] examined and compared the North American physical fiber network and the pan-European fiber optic network; Gorman and Malecki [16] analyzed the complexities of ten backbone provider networks in the USA; Wheeler and O'Kelly [17] and Grubesic et al. [18] explored the network topology of the commercial Internet in the USA. Such studies address the position of nodes (e.g., cities and hubs) as well as the diffusion pattern of flows in the network, with due emphasis on route length, nodal clustering, and power-law and exponential connectivity distributions.

In large-scale communication networks, such as the Internet, scale-free characteristics evolve through self-organizing processes, in which new nodes tend to attach to other vertices that are already well connected [19–22]. As a result, the connections in the network are mainly routed by several high-level nodes, and the network diameter is small, while network efficiency is high: Albert et al. [12] indicated that scale-free networks, such as these, are very tolerant to random failures at nodes—providing a reliable system for information distribution. However, Grubesic et al. [18] and Schintler et al. [15] pointed out that scale-free network, when super connected nodes are removed, faces the risk of disconnection or significant interruption, either unintentionally or from a targeted attack or external force.

However, so far, the reliability, survivability, and resiliency of a communication network system, within the context of complex network theory, are still ignored to a great extent, meaning the related literature is relatively scant, especially concerning the survivability of a submarine cable network. In the literature, a few scholars have focused on the topological design of submarine cable networks [23–25]. A disaster in the network is often considered as a randomly placed disk of a particular radius (the closer the layout of submarine cables in the disaster area, the more the cables will be affected). Computational geometric techniques are used to compare and evaluate the network resiliency for various topologies, such as rectangular with rounded corner, rhombus, and elliptic topologies, and to find the vulnerable points within a network. Some studies even involve the constraints of the total cable cost [26]. The above research focuses on the vulnerability or survivability of a submarine cable network; however, the essence is to explore the optimization of spatial layouts of a submarine cable network in the context of communication engineering. All the above research studies are not based on complexity science and lack the dynamic evaluation of network complexity. The work of Albert et al. [12] and that of other scholars on the tolerance of the Internet to random failures or targeted attacks need to be further reflected in a submarine cable network using complex network theory.

The first international submarine cable system, invested by China, was completed and came into use in December 1993. In the subsequent 25 years, the number of submarine cable systems worldwide has increased, the capacity of submarine cable systems has improved, the number of connected countries and regions has increased, and undersea communication capacity has reached new heights [27]. Nevertheless, similar to the characteristics of shipping routes, a few sea lanes or landing countries have become the necessary places for network connectivity, greatly restricting the connectivity of China's submarine cable network. However, with global warming, the melting Arctic has provided an attractive route for submarine cables [28]. In March 2017, at the 4th International Arctic Forum held in Arkhangelsk, Russia, many parties, including the Ministry of Industry and Information Technology of China and China Telecom Group, discussed the planning and construction of a submarine cable across the Arctic circle along the Northeast channel. Such a project can shorten the submarine cable connection distance of Asia and Europe by about 40% and reduce the communication delay by half. Against this background, the submarine cable network of China will usher in a new round of changes. How network vulnerability will change requires a scientific answer from the perspective of academic.

Thus, this paper draws on the complex network theory and method to analyze the vulnerability of the submarine cable network of mainland China and compares the vulnerability between before and after construction of the trans-Arctic cable system. The main features and contributions of this paper can be summarized as follows.

(1) This paper is the first attempt to construct the submarine cable network of Mainland China. It paper presents its structural characteristics and makes up for the deficiency of existing research on this new large-scale connectivity infrastructure network.

(2) This paper proposes a quantitative method to measure the vulnerability of submarine cable network. This method fully combines the perspective of topological network and geographical space, which can not only simulate the change trend of network eigenvalues under different attack modes but also identify the important nodes and corridors in the global map.

(3) This paper proves the strategic significance of the trans-Arctic cable system to the submarine cable network of mainland China. It shows that the vulnerability of the submarine cable network of mainland China will be improved after construction of the trans-Arctic cable system, especially in the context of intentional attack.

The remainder of this paper is organized as follows. In Section 2, the data sources and study area are presented; the submarine cable network of mainland China is constructed. In

Section 3, the methodology is proposed, including assessment indicators, simulation strategy, and quantitative analysis. In Section 4, results about the vulnerability analysis of submarine cable network nodes and edges are separately provided. The conclusion and discussion are finally drawn in Section 5.

2. Data Sources and Network Construction

2.1. Data Sources and Study Area. The data for this study mainly come from TeleGeography (https://www.telegeography.com/), which is a telecommunications market research and consulting firm. TeleGeography collected the data for all existing submarine communications cables. For this study, we analyze the submarine cable network of mainland China, which comprises part of the global submarine cable network and consists of all the submarine cable systems arriving in mainland China. Fourteen submarine cable systems were selected as the research object (Table 1). Among them, 13 submarine cable systems have been built—3 have been abandoned, and ten are currently in use. The TAC system is expected to be completed and put into use in 2023. It should be noted that although the AAE-1 system landed first in Ngwe Saung, Myanmar, it directly connects with China through the China-Myanmar international fiber-optic cable system (CMI project), and the service object is still mainland China (Figure 1). This system is also the first overseas submarine cable built by China Unicom with its own landing station; therefore, the AAE-1 system belongs to the research consortium of the submarine cable network of mainland China.

Submarine cable network of Mainland China connects 44 countries and regions. Before completion of the TAC system (later called the T_1 era), the submarine cable network of mainland China was connected to the Asia-Pacific, South Asia, the Middle East, North Africa, and Europe by the west route of China and connected to the United States by the east route of China. After completion of the TAC system (later called the T_2 era), the new route along the Arctic Northeast channel connects Europe directly to mainland China, forming a closed-loop structure. The main routes, countries, and regions involved are shown in Figure 1.

2.2. Construction of the Submarine Cable Network. In order to facilitate theoretical analysis, the submarine cable network is abstracted as nodes, edges, and weights. The node set ($V = \{v_i: i = 1, 2, \ldots, n\}$) is defined as the landing countries or regions (v_i); the edge set ($E = \{e_i: i = 1, 2, \ldots, m\}$) comprises the node pairs with optical cable lines directly between nodes (e_i); the weight of each edge ($W = \{w_{ij}: i = 1, 2, \ldots, n\}$) is determined by the capacity (potential capacity, rather than lit capacity, is used here because redundancy is necessary to avoid meltdown [29]) of the submarine cable system. The matrix expression is as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1j} \\ a_{21} & a_{22} & \ldots & a_{2j} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{ij} \end{bmatrix}. \quad (1)$$

According to equation (1), the submarine cable network of mainland China is constructed and visualized, as shown in Figure 2. The connection level of the submarine cable network of mainland China is characterized by clustering and distance attenuation. Countries or regions in East Asia have the highest connection level, and Southeast Asian groups the second-highest level. Europe, Africa, the Middle East, and South Asia have a low level of connection, while the United States has a medium level. After construction of the TAC system, Russia and northern Europe will also reach a medium level.

From preliminary observation, the network has complex connections and presents the characteristics of small world. The average shortest path and clustering coefficient of submarine cable of mainland China and same-size random network are calculated, respectively (Table 2). The results show that the clustering coefficient of submarine cable of mainland China is much larger than that of same-size random network, which indicates that submarine cable of mainland China has the characteristics of small world. In addition, the average shortest path of submarine cable of mainland China is slightly higher than that of same-size random network. Due to the limitation of marine geography, the submarine cable cannot be connected at will. On this basis, it is necessary to make a further analysis of its network vulnerability.

3. Methodology

3.1. Assessment Indicators. Network vulnerability refers to the extent to which network connectivity is affected when network structure is impacted. The vulnerability of a submarine cable network refers to the extent to which network connectivity is affected when several landing stations or submarine optical cable lines cannot operate normally due to random damage caused by natural factors, such as earthquake, tsunami, anchor damage, and shark bite or intentional destruction caused by unnatural factors, such as fishing vessel stealing, terrorism, and irregular war operations. Therefore, submarine cable network vulnerability can be analyzed using the changes in eigenvalues after the network is attacked.

To understand the impact of failures and attacks on network structure, we choose four mainstream indicators to measure network vulnerability when nodes are removed: the average degree of networks D, the clustering coefficient of the entire network C, the proportion of isolated nodes ΔN, and the global efficiency E.

3.1.1. Average Degree of Networks D. The average degree of networks D is the average value of all node degrees (the number of edges connected with nodes) in the network. When the network is attacked, with the corresponding decrease in network nodes and edges, the network average degree will inevitably change. The greater the change is, the more sensitive and vulnerable the network is, which can be expressed as follows:

Table 1: Major international submarine cable systems landing in Mainland China.

No.	System	Code	Bandwidth capacity	Length (km)	Completed time	Remarks
1	China-Japan fiber-optic submarine cable system	C-J	560 mbps	1252	Dec, 1993	Abandoned in 2006
2	China-Korea fiber-optic submarine cable system	C-K	1120 mbps	549	Feb, 1996	Abandoned in 2005
3	FLAG Europe Asia	FLAG	10 Gbps	27000	Sep, 1997	In service
4	South-East Asia-Middle East-West Europe 3	SMW3	960 Gbps	39000	1999	In service
5	China-US cable network	CUCN	80 Gbps	30800	Jan, 2000	Abandoned in 2016
6	Asia Pacific cable network 2	APCN2	2.56 Tbps	19000	Dec, 2000	In service
7	East Asia crossing	EAC	2.56 Tbps	19850	Jan, 2002	In service
8	City-to-City cable system	C2C	7.68 Tbps	17000	Aug, 2002	In service
9	Trans-Pacific Express	TPE	5.12 Tbps	17700	Sep, 2008	In service
10	South-East Asia Japan cable system	SJC	15 Tbps	10700	Feb, 2013	In service
11	Asia Pacific gateway	APG	54 Tbps	10400	2016	In service
12	Asia-Africa-Europe-1 cable system	AAE-1	40 Tbps	25000	2017	In service
13	New crossing-Pacific cable system	NCP	60 Tbps	13618	2018	In service
14	Trans-Arctic cable system	TAC	60 Tbps	12700	by 2023	Planned

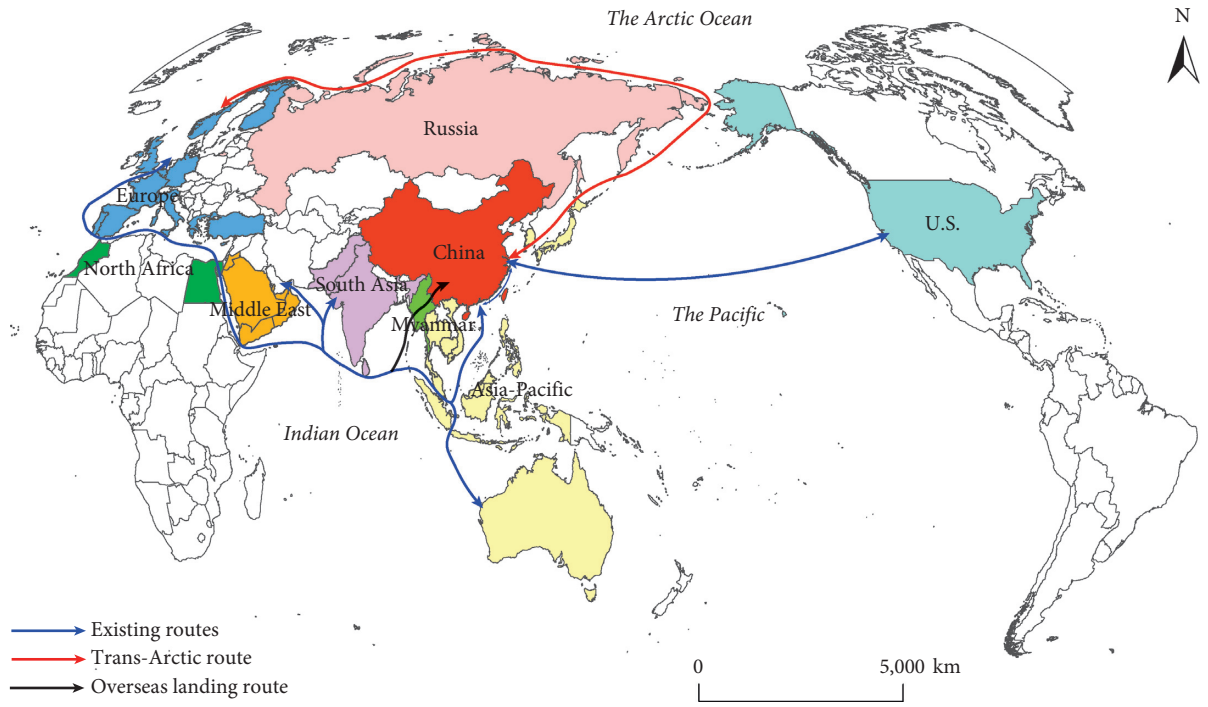


Figure 1: Main routes of the submarine cable network of mainland China.

$$D = \frac{1}{N}\sum_{i=1}^{N} D_i, \tag{2}$$

where N is the total number of nodes in the network, D_i is the degree value of node i, and D is the average degree of networks.

3.1.2. Clustering Coefficient of the Entire Network C. The clustering coefficient [30] reflects the connection between adjacent nodes in the network. The higher the clustering coefficient, the easier it is to form regional agglomeration between a point and its surrounding nodes; the lower the clustering coefficient, the more difficult it is to form regional agglomeration in the spatial distribution. In general, when the network is attacked, the clustering coefficient will decrease and network structure will become loose. This can be expressed as follows:

$$C_i = \frac{2M_i}{k_i(k_i - 1)}, i = 1, 2, 3, \ldots N, \tag{3}$$

where k_i is the number of nodes directly connected to node i, k_i $(k_i–1)/2$ is the maximum number of edges of node i connected pairwise, M_i is the number of edges between adjacent nodes of node i, and C_i represents the concentration level of the nodes.

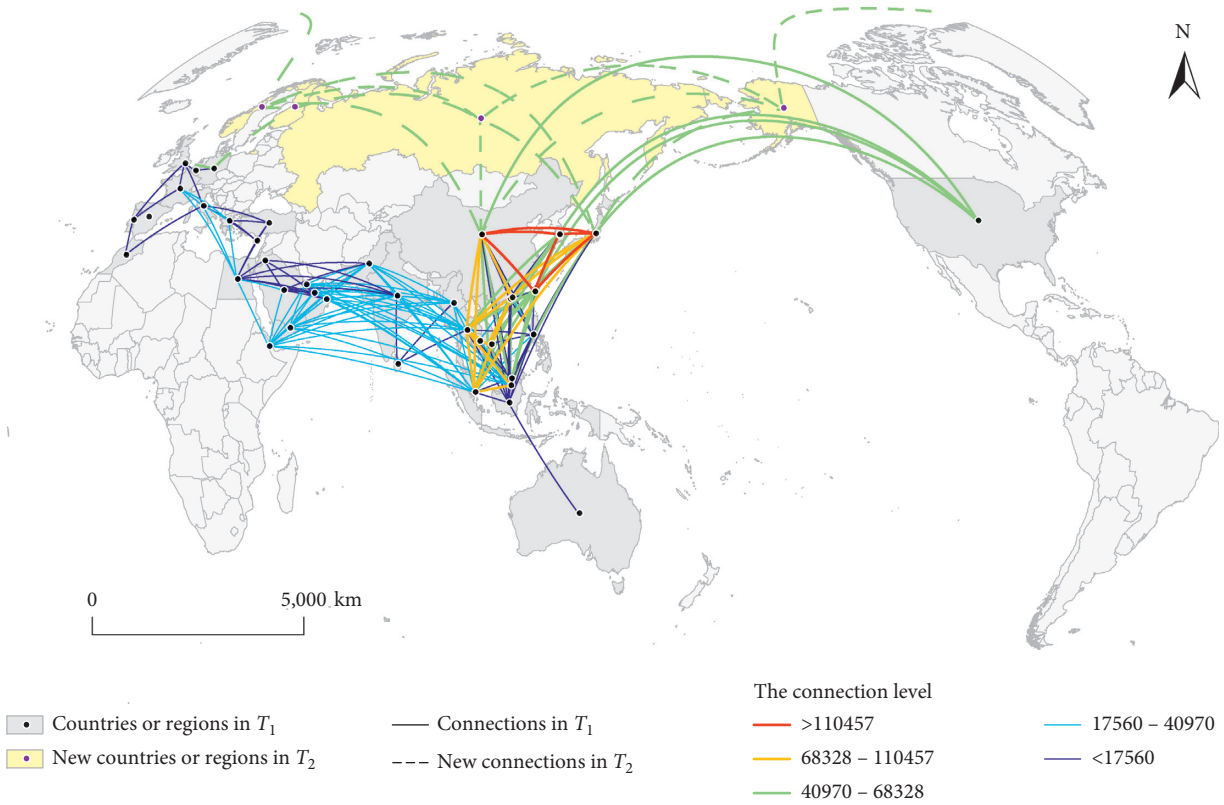


Figure 2: Connection of submarine cable network of mainland China.

Table 2: Comparison of submarine cable network of mainland China and same-size random network metrics.

	Number of nodes	Probability of linked edges	Average shortest path	Clustering coefficient
Submarine cable network of mainland China (T_1)	38	0.24	2.69	0.729
Random network (T_1)	38	0.24	1.844	0.237
Submarine cable network of mainland China (T_2)	42	0.21	2.649	0.727
Random network (T_2)	42	0.21	1.909	0.211

The clustering coefficient of the entire network C can then be obtained by calculating the average value of the clustering coefficient of each node:

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i. \tag{4}$$

3.1.3. Proportion of Isolated Nodes ΔN. The proportion of isolated nodes in a network is the proportion of nodes without edges connected. When the network is attacked, the edges between some nodes become disconnected, creating some isolated nodes and affecting the scale and connectivity of the whole network. This can be expressed as follows:

$$\Delta N = \left(1 - \frac{N'}{N}\right) \times 100\%, \tag{5}$$

where N and N' represent the total number of nodes before and after attack, respectively, and ΔN represents the proportion of isolated nodes.

3.1.4. Global Efficiency E. The global efficiency E reflects the difficulty of information transmission between nodes in the network, defined by Latora and Marchiori [31]. The higher the network efficiency, the better the network connectivity. The efficiency between node i and node j in the network is the reciprocal of the shortest distance d_{ij} between the two nodes. When i and j are unconnected, $d_{ij} = +\infty$, so $1/d_{ij} = 0$. For the whole network, the mean value of efficiency between all nodes is the global efficiency, represented by E:

$$E = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{d_{ij}}, \quad i \neq j. \tag{6}$$

3.2. Simulation Strategy

3.2.1. Node Attack. Our approach to analyze the network vulnerability involved conducting successive simulated attacks on the network to test its performance until it collapsed. The faults of submarine cable networks can be divided into an arc failure (the severing of an optical cable) and node failure (the failure of landing station equipment) [32]; the modes of failure can be divided into intentional attack and random attack.

Random attack: network nodes are attacked randomly with a certain probability; that is, the nodes in the network are randomly deleted according to the probability. Such a random selection is not related to the topological characteristics or any other attributes of a node. This attack method can simulate the impact of earthquake, tsunami, anchor damage, shark bite, and other random events, on the submarine cable network.

Intentional attack: network nodes with great influence sustain a targeted attack; that is, the nodes in the network are deleted according to node importance. This mode of attack can simulate the impact of intentional events, such as fishing vessel theft, terrorism, and military blockade, on the submarine cable network.

The routes in the submarine cable network of mainland China are concentrated in a minority of hub countries or regions. Due to differences in the capacity of optical fiber systems, the importance of network nodes with the same connection but different capacities is different; it is therefore necessary to weight the degree value of each node in the network, ranking the importance of nodes in the submarine cable network of mainland China. Among them, the top ten node countries or regions are shown in Table 3.

In order to reveal the vulnerability of the network more clearly, we gradually delete nodes to simulate random attack and intentional attack, respectively. We then evaluate the changes in the average degree of networks D, the clustering coefficient of the entire network C, the proportion of the isolated nodes ΔN, and the global efficiency E.

3.2.2. Edge Attack. Due to limitations in marine geography, the submarine cable lines must to be laid through channels or canals. Therefore, the security of these lanes has a direct impact on the vulnerability of the submarine cable network. The submarine cable of mainland China passes through eight main sea lanes (Figure 3; Table 4). As these sea lanes are accident-prone areas with a low security guarantee [32], they are important globally, strategically, and militarily. Therefore, it is important to analyze the impact of these sea lanes on the vulnerability of the submarine cable network.

The attack on each sea lane is simulated. When attacking, we break the edges and remove the isolated nodes, thus forming a new submarine cable network under attack. Based on this, each eigenvalue of the network is calculated and compared with the normal eigenvalue of the network so as to evaluate the vulnerability of the submarine cable network under edge attack.

3.3. Quantitative Analysis

3.3.1. Network Stress Testing. To quantitatively analyze the change degree of network vulnerability, the key is to determine the contribution degree of each network eigenvalue rate of change to network vulnerability—that is, the corresponding weight.

Therefore, we use the network stress test method, and the related concepts are as follows:

(i) *Network Half-Attenuation Degree G_s.* The network half-attenuation degree G_s represents the attack intensity when the rate of change in a given eigenvalue reaches 50% of its maximum value (that is, the network completely fails); s is the given eigenvalue. Generally, it is best to conduct stress testing at a ratio of 50%—that is, the median value of network integrity and complete failure. Therefore, we use the network half-attenuation degree G_s to determine the sensitivity coefficient.

(ii) *Sensitivity Coefficient O_s.* The sensitivity coefficient O_s is the ratio of the half-attenuation degree G_s of the network in T_2 to the half-attenuation degree G_s of the network in T_1, expressed as follows:

$$O_s = \frac{G_S(T_2)}{G_S(T_1)}, \quad s = 1, 2, \ldots. \tag{7}$$

(iii) *Weight of Impact Q_s.* The weight of impact Q_s refers to the contribution of a given eigenvalue to network vulnerability, expressed as follows:

$$Q_s = \frac{O_s}{\sum_s O_s}, \quad s = 1, 2, \ldots. \tag{8}$$

3.3.2. Variation Degree of Network Vulnerability. We define the variation degree of network vulnerability, which represents the value of change in network vulnerability before and after interconnection of the TAC system.

Let $U_{s,\ h}$ be the set of the rate of change in network eigenvalue s when the attack ratio is h. For example, $U_{1,\ 1\%}$ represents the rate of change in the network average degree when the attack ratio is 1%. Let $\overline{U_s}(T_t)$ be the average rate of change in a given eigenvalue s of the network under different attack ratios; then:

$$\overline{U_s}(T_t) = \frac{\sum_h U_{s,h}}{\mathrm{num}(h)}, \quad t = 1, 2; s = 1, 2, \ldots; h = 1\%, 2\%, \ldots. \tag{9}$$

Let $\Delta\overline{U_s}$ be the difference in the average rate of change in eigenvalues at T_1 and T_2; then:

$$\Delta\overline{U_s} = \overline{U_s}(T_2) - \overline{U_s}(T_1), \quad s = 1, 2, \ldots. \tag{10}$$

Let F be the variation degree of network vulnerability; then:

$$F = \sum_s [Q_s \times \Delta\overline{U_s}], \quad s = 1, 2, \ldots. \tag{11}$$

Table 3: Top ten countries or regions with the largest weighted degree.

Rank	T_1		T_2	
	Country or region	Weighted degree	Country or region	Weighted degree
1	Thailand	1060860	Mainland China	1080700
2	Singapore	970920	Japan	1069170
3	Malaysia	843860	Thailand	1060860
4	Mainland China	840700	Singapore	970920
5	Japan	829170	Malaysia	843860
6	Taiwan, China	729600	Taiwan, China	729600
7	Korea	727070	Korea	727070
8	Hong Kong, China	701830	Hong Kong, China	701830
9	Vietnam	557760	Vietnam	557760
10	Djibouti	485760	Djibouti	485760

Figure 3: Location of main sea lanes (see Table 4 for the meaning of the numbers in Figure 3).

Table 4: Main sea lanes.

No.	Sea lanes
1	Taiwan Strait
2	Strait of Malacca
3	Bab-el-Mandeb
4	The Suez Canal
5	Strait of Gibraltar
6	Strait of Hormuz
7	The Pacific
8	Bering Strait

4. Results

4.1. Vulnerability Analysis of Submarine Cable Network Nodes. In order to test the changes of each assessment indicator, a number of simulations under different attack scenarios have been performed through MATLAB 2019b, utilizing a personal computer with intel Core i7-8550U CPU @ 1.80 GHz and 8 GB of RAM. Different deletion strategies for network nodes are considered in Section 3.2.1. *Node attack*: to ensure the data stability in random attack mode, *numRandom* is set as 500. Considering the different network scale in different eras, the parameters of deleting nodes should be different, and thus *numdelete* is set as 38 and 42, respectively, in T_1 and T_2. Simulation results of changes in each network eigenvalues are shown in Figure 4.

4.1.1. Changes in the Eigenvalues. A comparison of the two different attack modes shows that under a random attack, D, C, and E decrease slowly with increasing attack proportion, while ΔN increases slowly. In the case of intentional attack, the above four eigenvalues change more significantly. When the attack ratio reaches a certain level, D, C, and E show a sharp decreasing trend, reaching 0, and ΔN increases sharply, reaching 1. These results show that the submarine cable network is relatively robust under a random attack but relatively vulnerable under an intentional attacked.

The same attack mode in different periods is then compared. In T_1 and T_2, the overall trends in network eigenvalue change are nearly the same in either attack mode. This indicates that construction of the TAC does not cause a dramatic change in the submarine cable networks of mainland China, but an additional supplement. Moreover, at the same proportion of attack nodes, there are slight or significant differences in the ranges of variation in network eigenvalues between the two periods. Although the ranges of variation in network eigenvalues in T_2 are slightly larger than that in T_1 for some intervals, the trend in variation of each eigenvalue of the submarine cable network in T_2 slightly lags behind that in T_1 (the specific analysis of each eigenvalue will be presented below). This indicates that the sensitivity of the submarine cable network to node attack is lower in T_2 than in T_1, which proves that construction of the TAC will reduce the vulnerability of the submarine cable network in mainland China.

In the following, we analyze the detailed characteristics of each eigenvalue change.

(i) D: the initial value of D in T_2 (8.32) is slightly smaller than that in T_1 (8.5), which is due to the addition of a few new nodes, such as Russia, Finland, and Norway. These new nodes are basically isolated from the other nodes in the original network, so D decreases slightly with increasing network scale. This is also the reason why the downward trend in D for intentional attack mode is not monotonous (the same principle applies to C). When the proportion of nodes deleted reaches about 80%, the D value for intentional attack being higher than that for random attack can also be explained by this. We compare the curves of the same attack mode for the two eras, and we find the following. (1) In random attack mode, 80% of the deleted nodes is a significant dividing line—before that, the trends of D in T_1 and T_2 are almost the same, but after that, D in T_1 drops sharply to 0. In T_2, when the proportion of deleted nodes reaches 84%, D drops abruptly, lagging 4% behind D in T_1. (2) In intentional attack mode, with increasing proportion of deleted nodes, the trend of D in T_1 and T_2 is uneven. When the proportion of deleted nodes is between 0–1% and 35–68%, D in T_2 is higher than D in T_1; when the proportion is between 1–35% and 68–82%, D in T_1 is higher than D in T_2. However, when D drops to 0, the proportion reaches 86% in T_2, lagging 4% behind T_1. The above shows that although the range of change in D in T_2 is larger in some intervals, the change trend of D in T_2 still lags behind that in T_1, whether random attack or intentional attack is considered.

(ii) C: firstly, it is clear that the descent speed of C under an intentional attack is significantly faster than that under a random attack, and the proportion of deleted nodes is 61% (T_1) and 67% (T_2), respectively, when C to drop to 0. However, in random attack mode, the proportion is 78% (T_1) and 82% (T_2), respectively. This proves that the submarine cable network in mainland China better resists random attack compared with intentional attack, under which it is relatively vulnerable. Moreover, under the same attack mode, the increase extent of the ratio of minimum deleted nodes with a C value of 0 for intentional attack (6%) is slightly higher than that for random attack (4%), which seems to indicate that the improvement in the response to intentional attack is more significant in T_2. From the details, in random attack mode, the change in C in T_2 is not always better than that in T_1. When the proportion of deleted nodes is less than 55%, the change in C in T_2 is similar to that in T_1, but when the proportion is between 55% and 78%, the decline speed of C in T_2 is faster. In intentional attack mode, the curves of T_1 and T_2 have similar peaks and troughs, but T_2 almost lags behind T_1.

(iii) ΔN: when the network is attacked, the number of isolated nodes will gradually increase, so ΔN shows an upward trend. However, the change in trend of ΔN under an intentional attack is obviously faster than that under a random attack. With regard to change characteristics, the rising trend in ΔN is basically the same for the two periods of random attack mode. In T_2, there is a slight lag phenomenon before 80% of the nodes are deleted. In intentional attack mode, the change in ΔN for T_1 and T_2 shows a characteristic of step-by-step increase. When the proportions of deleted nodes are 0–30% and 45–60%, ΔN in T_2 is lower than that in T_1; when the

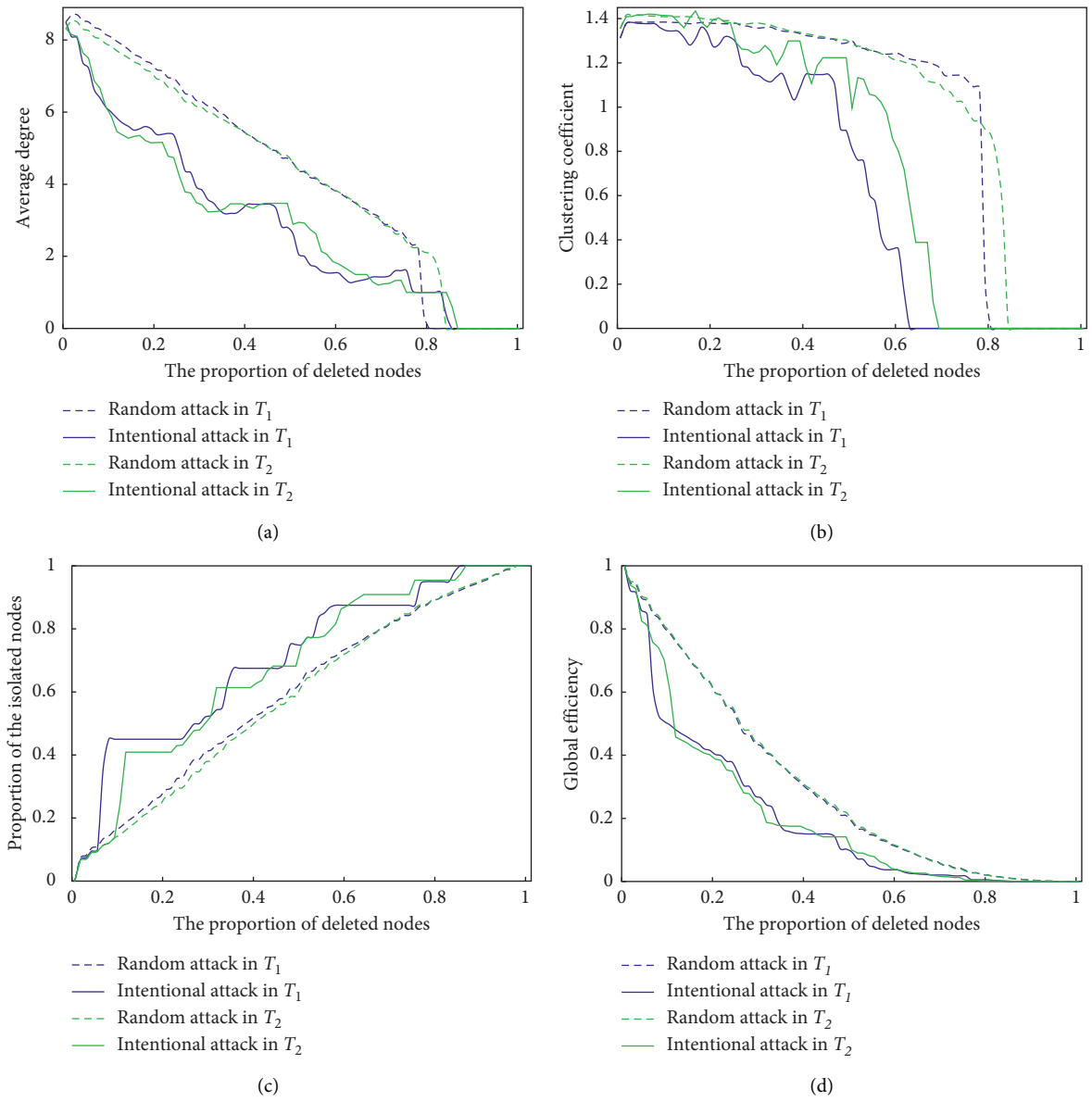


Figure 4: Changes in the network eigenvalues.

proportions are 30–45% and 60–82%, ΔN in T_1 is lower than that in T_2. In T_1, 82% of the nodes are deleted when ΔN decreases to 0; in T_2, the proportion of deleted nodes is 86%, which lags behind by 4%. This change is similar to that of D, which indicates that intentional attack has a greater impact on the submarine cable network in some intervals, but on the whole, the change in trend of ΔN in T_2 lags behind that in T_1.

(iv) E: in contrast with D and C, whether in random attack or intentional attack mode, the descending trend of E is at first fast and then slow. In random attack mode, the trend in E is basically the same for T_1 and T_2. When the proportion of deleted nodes reaches 95%, E tends to 0. In intentional attack mode, the curve of E for T_1 and T_2 traces a staggered downward trend. When the proportion of deleted nodes reaches about 85%, E drops to 0. When the proportion is within the range of 5–10%, E decreases fastest in T_1 and T_2, but the curve in T_2 obviously lags behind that in T_1. This proves that the construction of TAC can improve the ability of the submarine cable network of mainland China to deal with intentional attack under a certain proportion of deleted nodes.

Through the above analysis, a few hub landing stations in the submarine cable network of mainland China are of great importance; therefore, the network is relatively vulnerable under an intentional attack and relatively strong under a random attack. However, by comparing the changes in four eigenvalues in T_1 and T_2, we find that the influence of the submarine cable network of mainland China in T_2 lags

behind that in T_1, whether from the perspective of the critical point of network collapse or from the perspective of the change characteristics of each eigenvalue. This indicates that the construction of TAC reduces the vulnerability of the submarine cable network of mainland China to some extent, enhancing its ability to deal with attacks. Moreover, from the simulation results, the construction of TAC seems to have a different response effect on different attack modes—this is the goal of quantitative analysis in the following.

4.1.2. Quantitative Analysis. According to the analysis method in Section 3.3, we quantitatively analyze the vulnerability of the submarine cable network of mainland China under different attack modes. The calculation results are shown in Tables 5 and 6.

In random attack mode, $\Delta\overline{U_s}$ of D, C, ΔN, and E is −1.219%, 0.416%, −1.256%, and −0.406%, respectively. Among them, $\Delta\overline{U_s}$ of C is positive, indicating that this index is more vulnerable in T_2 than in T_1, while other network eigenvalues are negative, indicating that it is more vulnerable in T_1 than in T_2; Q_s of each eigenvalue is 0.248, 0.252, 0.252, and 0.248, respectively, and the difference is small. The contribution of each network characteristic to network vulnerability is −0.302%, 0.105%, −0.317%, and −0.101%, respectively. After the accumulation of all the contribution values, the variation degree of network vulnerability F is -0.615%. This value indicates that in random attack mode, the submarine cable network of mainland China in T_1 is more vulnerable than that in T_2, and the construction of TAC reduces network vulnerability.

In intentional attack mode, $\Delta\overline{U_s}$ of D, C, ΔN, and E is −0.983%, −7.538%, −2.236%, and −0.606%, respectively—all negative, which indicates that the submarine cable network of mainland China in T_1 is more vulnerable than that in T_2; Q_s of each eigenvalue is 0.213, 0.269, 0.252, and 0.266, respectively. The contribution of each network characteristic to network vulnerability is −0.209%, −2.025%, −0.564%, and −0.161%, respectively. After the accumulation of all the contribution values, the variation degree of network vulnerability F is −2.960%. This value indicates that in intentional attack mode, the submarine cable network of mainland China in T_1 is more vulnerable than that in T_2, and the construction of TAC reduces network vulnerability.

The results show that the submarine cable network of mainland China is more vulnerable now than after the construction of TAC, and the construction of TAC reduces network vulnerability to some extent. However, for different attack modes, the reduction of network vulnerability in intentional attack mode is more significant than that in random attack mode.

4.2. Vulnerability Analysis of Submarine Cable Network Edges. The curves of D, C, ΔN, and E are almost monotonous, directly reflecting the degree to which the network is affected by attack. Therefore, the above four network eigenvalues can be used to analyze the vulnerability changes in the network when the sea channels, that is, the edges in the network, are attacked.

4.2.1. Segmentation of Network by Edge Fracture. The submarine cable network of mainland China will be split in different ways due to the obstruction of different sea lanes. Table 7 lists the impacts of sea lane interruption on the submarine cable network of mainland China. The Taiwan Strait impacts the EAC and SMW3 systems, but it is not the sea lane through which the main cable passes, so the overall network structure is almost unchanged. The Strait of Malacca, Bab-el-Mandeb, the Suez Canal, and the Strait of Gibraltar are important sea lanes for mainland China to South Asia, the Middle East, Africa, and Europe. An attack on these sea lanes will break the network in T_1, meaning some areas become isolated subnetworks separated from mainland China. However, in T_2, an attack on these sea lanes will only cause some failures in the network, but the affected countries or regions still maintain contact with mainland China. The Strait of Hormuz is a branch lane, which only plays a key role in the branch route of the AAE-1 system—its breakage will affect the links between some Middle East countries and mainland China. The Pacific Ocean is an important sea lane of the submarine cable between China and the United States, and its obstruction will directly cause the United States to become an isolated subnet, no matter in T_1 or T_2. The Bering Strait becomes the only route which must be passed through in the TAC system in T_2—its breakage will cause Russia and other countries or regions to become separated from mainland China.

4.2.2. Importance Ranking of Sea Lanes. According to the simulation strategy in Section 3.2.2, this paper calculates the changes in network eigenvalues under the condition of each sea lane breaking so as to quantify and rank the importance of theses sea lanes (Table 8).

The results show that the Strait of Malacca is the most important sea lane in the submarine cable network of mainland China, both in T_1 and T_2. The average rate of change in network eigenvalues is highest when the Strait of Malacca is blocked. Bab-el-Mandeb and the Suez Canal rank second and third in importance in the network. The rankings for the Strait of Gibraltar, the Taiwan Strait, and the Pacific are relatively low, indicating a relatively low impact on the submarine cable network of mainland China. The Bering Strait becomes a new sea lane after completion of the TAC, with an average rate of change of 0.047, ranking fifth—its importance is less than that of the Strait of Hormuz.

However, from a comparison of the different periods, the average rate of change of the Malacca Strait, Bab-el-Mandeb, and the Suez Canal, ranked the top three in importance, decreases in T_2. Judging from the standard deviation of the changes in the eigenvalues, T_2 (0.083) is smaller than that of T_1 (0.124), which indicates that TAC construction reduces the importance of the Malacca Strait, Bab-el-Mandeb, and the Suez Canal in the submarine cable network of mainland China, and the difference of the importance of each sea lane in the whole network is reduced. From a geographical perspective, the Malacca Strait, Bab-el-Mandeb, and the Suez Canal are the only sea lanes connecting mainland China and the Middle East, and Africa and Europe, in T_1.

Table 5: Average rate of change in network eigenvalues.

Eigenvalues	Random attack			Intentional attack		
	$\overline{U_S}$ ($T1$)	$\overline{U_S}$ ($T2$)	$\Delta\ \overline{U_S}$	$\overline{U_S}$ ($T1$)	$\overline{U_S}$ ($T2$)	$\Delta\ \overline{U_S}$
D	50.185	48.966	−1.219	66.932	65.949	−0.983
C	23.360	23.776	0.416	50.140	42.602	−7.538
ΔN	60.380	59.123	−1.256	71.813	69.576	−2.236
E	71.565	71.159	−0.406	81.649	81.043	−0.606

Table 6: Analysis of variation degree of network vulnerability.

	Eigenvalues	Gs (T_1)	Gs (T_2)	Os	Qs	$\Delta\ \overline{U_S}$ (%)	F (%)
Random attack	D	53.731	55.309	1.029	0.248	−1.219	−0.302
	C	78.494	82.079	1.046	0.252	0.416	0.105
	ΔN	37.683	39.460	1.047	0.252	−1.256	−0.317
	E	25.363	26.093	1.029	0.248	−0.406	−0.101
	Total	—	—	5.144	1	—	**−0.615**
Intentional attack	D	27.860	25.258	0.907	0.213	−0.983	−0.209
	C	53.851	61.644	1.145	0.269	−7.538	−2.025
	ΔN	27.500	29.545	1.074	0.252	−2.236	−0.564
	E	9.666	10.970	1.135	0.266	−0.606	−0.161
	Total	—	—	4.261	1	—	**−2.960**

Table 7: Impact of sea lane interruption on the submarine cable network of mainland China.

No.	Sea lanes	Affected cable system	Impact on submarine cable network of mainland China	
			T_1	T_2
1	Taiwan Strait	EAC, SMW3	A few areas are affected, but the overall network structure is almost unchanged.	
2	Strait of Malacca	FLAG, SMW3, AAE-1	South Asia, the Middle East, Africa, and Europe become isolated subnetworks, separated from mainland China.	The network is partly broken, but the affected countries or regions can still maintain contact with mainland China.
3	Bab-el-Mandeb	FLAG, SMW3, AAE-1	Africa, Europe, and parts of the Middle East become isolated subnetworks, separated from mainland China.	
4	The Suez Canal	FLAG, SMW3, AAE-1	Africa and Europe become isolated subnetworks, separated from mainland China.	
5	Strait of Gibraltar	FLAG, SMW3	Parts of Europe become isolated subnetworks, separated from mainland China.	
6	Strait of Hormuz	AAE-1	Parts of the Middle East become isolated subnetworks, separated from mainland China.	
7	The Pacific	TPE, NCP,	The U.S. mainland becomes isolated subnetwork, separated from mainland China.	
8	Bering Strait	TAC	None.	Russia becomes separated from mainland China.

Table 8: Changes in the eigenvalues when sea lanes are interrupted.

Era	No.	Sea lanes	D		ΔN	C		E		Average rate of change	Rank
			Value	Rate	Rate	Value	Rate	Value	Rate		
T1	1	Taiwan Strait	8.895	0.006	0.000	0.730	−0.001	0.240	0.006	0.003	7
	2	Strait of Malacca	4.158	0.535	0.632	0.836	−0.147	0.112	0.535	0.389	1
	3	Bab-el-Mandeb	7.053	0.212	0.316	0.743	−0.019	0.191	0.212	0.180	2
	4	The Suez Canal	7.842	0.124	0.263	0.720	0.012	0.212	0.124	0.131	3
	5	Strait of Gibraltar	8.579	0.041	0.105	0.708	0.029	0.232	0.041	0.054	4
	6	Strait of Hormuz	8.421	0.059	0.026	0.697	0.044	0.228	0.059	0.047	5
	7	The Pacific	8.737	0.023	0.026	0.711	0.025	0.236	0.024	0.025	6
	—	Normal	8.947	—	—	0.729	—	0.242	—	—	—

Table 8: Continued.

Era	No.	Sea lanes	D		ΔN	C		E		Average rate of change	Rank
			Value	Rate	Rate	Value	Rate	Value	Rate		
	1	Taiwan Strait	8.381	0.038	0.048	0.694	0.045	0.204	0.038	0.034	6
	2	Strait of Malacca	4.333	0.503	0.500	0.768	−0.056	0.106	0.503	0.290	1
	3	Bab-el-Mandeb	6.952	0.202	0.286	0.795	−0.094	0.170	0.202	0.119	2
	4	The Suez Canal	7.667	0.120	0.214	0.773	−0.063	0.187	0.120	0.078	3
T2	5	Strait of Gibraltar	8.190	0.060	0.095	0.776	−0.067	0.200	0.060	0.030	7
	6	Strait of Hormuz	7.952	0.087	0.071	0.685	0.058	0.194	0.087	0.061	4
	7	The Pacific	8.524	0.022	0.024	0.689	0.052	0.208	0.022	0.024	8
	8	Bering Strait	8.095	0.071	0.095	0.729	−0.003	0.197	0.071	0.047	5
	—	Normal	8.714	—	—	0.727	—	0.213	—	—	—

However, after completion of the TAC system, the route through the Bering Strait becomes a new choice. This is conducive to a reduction in network vulnerability.

5. Conclusion and Discussion

This paper has taken a first step in analyzing the vulnerability of the submarine cable network of mainland China, comparing changes in vulnerability before and after construction of the TAC system, from both complex network and geographical perspectives. The results can be summarized as follows.

The submarine cable network of mainland China is more robust under a random attack but more vulnerable under an intentional one, just like any other transport network with complex characteristics [33]. This is because the connections in the network are mainly routed through high-level node countries (such as Thailand, Singapore, Japan) and significant sea lanes (such as the Strait of Malacca and the Suez Canal). Intentional attacks carried out on these nodes and edges will readily lead to a collapse of the submarine cable network of mainland China, especially with respect to the connection between China and European countries, which relies almost exclusively on a single route from east to west.

However, if the TAC is built, this problem will be alleviated. Whether under a random attack or an intentional attack, the change in trend of each network eigenvalue will lag behind the current change, and the comprehensive score of the vulnerability change will be reduced, proving that the TAC system can improve the robustness of the submarine cable network of mainland China. Notably, the variation in the degree of network vulnerability under an intentional attack decreases more significantly than under a random attack. This indicates that the TAC system helps to enhance the performance of the submarine cable network of mainland China, when under an intentional attack. In fact, the TAC system not only strengthens the existing network structure but also enriches it.

Although the Strait of Malacca and the Suez Canal always possess critical strategic significance, the TAC system will weaken their importance. Judging by the change in trend of F value, the downward trend is not significant; however, the TAC system will still become an important project to change the situation of China and even the world submarine cable network. From a geographical perspective, the TAC system provides a new connection path; from the perspective of topology, the TAC system changes the large-chain structure of the network into a large-ring structure.

Our results have important implications for network vulnerability, in intentional attack mode, which should be taken into account by managers who analyze and assess the communication network security. Historically, submarine cables usually have been vulnerable to damage from threats related to fisheries, anchors, earthquakes, and tsunamis, but the increasingly assertive foreign policies of some states mean that these cables are at risk of purposeful interference, as per the emerging "hybrid" security threat. For example, Russia cut the submarine cables between Crimea and Ukraine in 2013—a crucial step in controlling the Internet within the annexed territory [34]. Therefore, in the information age, it is necessary for managers to have a good knowledge of vulnerable nodes and lines of submarine cable networks. This paper makes up for the lack of research in this field, as well as proving the importance of TAC construction in the context of the submarine cable network of mainland China.

However, there are some limitations in this paper, which could be addressed in future work. Although we fully consider the scenario of submarine cable systems having multiple landing stations in different cities in the same country, we still regard the country rather than the city as the network node in the network attack simulation. This will not affect the result of network vulnerability, but it is necessary to study from the microperspective, which has been done in previous studies [26]. In this work, the submarine cable network of mainland China is regarded as an independent local-area network and is thus not affected by the global submarine cable network. However, the submarine cable network of mainland China is not an isolated system and has various relationships with other submarine cables and regions [35]. Therefore, details may be lost when building the network, and research on the global submarine cable network will be a goal for future work. Moreover, event detection, as the primary strategy to deal with submarine cable network attacks, should be the future direction of submarine cable network vulnerability research. Especially, the graph-theory-based network partitioning algorithm has been applied to power system [36], which has important enlightenment to submarine cable transmission system.

Data Availability

The data for this study mainly come from TeleGeography (https://www.telegeography.com/), which is a telecommunications market research and consulting firm.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (no. 42071151) and the Strategic Priority Research Program of the Chinese Academy of Sciences (no. XDA20010101).

References

[1] V. Nagpal, *Convergence of Subsea Fiber, Terrestrial Fiber and Data Centers Leading to the Continental Edge and Subsea Colocation*, Offshore energy, Teesside, UK, 2019, https://suboptic2019.com/suboptic-2019-papers-archive/.

[2] H. Nakamoto, A. Sugiyama, and A. Utsumi, "Submarine optical communications system providing global communications network," *Fujitsu Scientific and Technical Journal*, vol. 45, pp. 386–391, 2009.

[3] D. A. Boughton and E. R. O. Smith, "Regional vulnerability: a conceptual framework," *Ecosystem Health*, no. 5, pp. 312–322, 1999.

[4] N. Adger and N. Kelly, "Social vulnerability to climate change and the architecture of entitlements," *Mitigation and Adaptation Strategies for Global Change*, no. 4, pp. 253–266, 1999.

[5] Y. Tian and H. Chang, "Bibliometric analysis of research progress on ecological vulnerability in China," *Acta Geographica Sinica*, vol. 67, no. 11, pp. 1515–1525, 2012.

[6] S. L. Cutter, B. J. Boruff, and W. L. Shirley, "Social vulnerability to environmental hazards ∗," *Social Science Quarterly*, vol. 84, no. 2, pp. 242–261, 2003.

[7] Yi Liu, J. Huang, and Li Ma, "The assessment of regional vulnerability to natural disasters in China based on DEA model," *Geographical Research*, vol. 29, no. 07, pp. 1153–1162, 2010.

[8] L. Adrianto and Y. Matsuda, "Developing economic vulnerability indices of environmental disasters in small island regions," *Environmental Impact Assessment Review*, vol. 22, no. 4, pp. 393–414, 2002.

[9] C. Fang, Y. Wang, and J. Fang, "A comprehensive assessment of urban vulnerability and its spatial differentiation in China," *Journal of Geographical Sciences*, vol. 26, no. 2, pp. 153–170, 2016.

[10] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[11] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[12] R. Albert, H. Jeong, and A. L. Barabasi, "Attack and error tolerance in complex networks," *Nature*, vol. 406, no. 6794, pp. 387–482, 2000.

[13] A. Broder, R. Kumar, F. Maghoul et al., "Graph structure in the web," *Computer Networks*, vol. 33, no. 1-6, pp. 309–320, 2000.

[14] B. . Bollobás and O. Riordan, "Robustness and vulnerability of scale-free random graphs," *Internet Mathematics*, vol. 1, no. 1, pp. 1–35, 2003.

[15] L. A. Schintler, S. P. Gorman, A. Reggiani et al., "Complex network phenomena in telecommunication systems," *Networks and Spatial Economics*, vol. 5, no. 4, pp. 351–370, 2005.

[16] S. P. Gorman and E. J. Malecki, "The networks of the Internet: an analysis of provider networks in the USA," *Telecommunications Policy*, vol. 24, no. 2, pp. 113–134, 2000.

[17] D. C. Wheeler and M. E. O'Kelly, "Network topology and city accessibility of the commercial Internet," *The Professional Geographer*, vol. 51, no. 3, pp. 327–339, 1999.

[18] T. H. Grubesic, M. E. O'Kelly, and A. T. Murray, "A geographic perspective on commercial Internet survivability," *Telematics and Informatics*, vol. 20, no. 1, pp. 51–69, 2003.

[19] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelley, "Identifying, understanding, and analyzing critical infrastructure interdependencies," *Control Systems Magazine*, vol. 21, no. 6, pp. 11–25, 2001.

[20] E. J. Malecki, "The economic geography of the internet's infrastructure," *Economic Geography*, vol. 78, no. 4, pp. 399–424, 2002.

[21] M. Barthélemy, "Crossover from scale-free to spatial networks," *Europhysics Letters (EPL)*, vol. 63, no. 6, pp. 915–921, 2003.

[22] S. P. Gorman and R. Kulkarni, "Spatial small worlds: new geographic patterns for an information economy," *Environment and Planning B: Planning and Design*, vol. 31, no. 2, pp. 273–296, 2004.

[23] P. K. Agarwal, A. Efrat, S. K. Ganjugunte et al., "Network Vulnerability to Single, Multiple, and Probabilistic Physical attacks," in *Proceedings of the MILICOM, Military Communication Conference*, San Jose, CA, USA, November 2010.

[24] S. Neumayer and E. Modiano, "Network reliability under random circular cuts," in *Proceedings of the IEEE GLOBECOM 2011*, pp. 1–6, Houston, Texas, USA, December 2011.

[25] L. M. Dawson, D. Ferhat, Z. Moshe et al., "Disaster-aware submarine fiber-optic cable deployment for mesh networks," *Journal Lightwave Technol*, vol. 34, pp. 4293–4303, 2016.

[26] C. Cao, M. Zukerman, W. Wu, J. H. Manton, and B. Moran, "Survivable topology design of submarine networks," *Journal of Lightwave Technology*, vol. 31, no. 5, pp. 715–730, 2013.

[27] Y. Ye, X. Jiang, G. Pan et al., *Submarine Optical Cable Engineering*, pp. 59–86, Academic Press, Cambridge, UK, 2018.

[28] J. Saunavaara and M. Salminen, "Geography of the global submarine fiber-optic cable network: the case for arctic ocean solutions," *Geographical Review*, 2020.

[29] J. Holt and P. Vonderau, "'Where the internet lives': data centers as cloud infrastructure," in *Signal Traffic: Critical Studies of Media Infrastructures*, L. Parks and N. Starosielski, Eds., pp. 71–93, University of Illinois Press, Champaign, IL, USA, 2015.

[30] O. Woolley-Meza, C. Thiemann, D. Grady et al., "Complexity in human transportation networks: a comparative analysis of worldwide air transportation and global cargo-ship movements," *The European Physical Journal B*, vol. 84, no. 4, pp. 589–600, 2011.

[31] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, no. 19, p. 198701, 2001.

[32] A. Lisser, R. Sarkissian, and J. P. Vial, *Survivability in Telecommunication Networks. Manuscript, HEC, Section of*

Management Studies, University of Geneva, Geneva, Switzerland, 1995.

[33] J. Lu and T. Gao, "Efficiency of safety control in key nodes of international sea lanes," *China Soft Scince*, no. 10, pp. 1–8, 2015.

[34] L. Robert and P. Stephen, *Little Green Men: A Primer on Modern Russian Unconventional Warfare*, Ukraine 2013-2014, Ukraine, European, 2019.

[35] Telegeography, "The submarine cable map," *Global Bandwidth Research Service*, Telegeography, Washington, DC, USA, 2016, http://www.submarinecablemap.com/.

[36] D. Ma, X. Hu, H. Zhang, Q. Sun, and X. Xie, "A hierarchical event detection method based on spectral theory of multidimensional matrix for power system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 99, pp. 1–14, 2019.

Hindawi
Complexity
Volume 2021, Article ID 6641030, 15 pages
https://doi.org/10.1155/2021/6641030

Research Article

Node Importance Evaluation of Cyber-Physical System under Cyber-Attacks Spreading

Xin-Rui Liu, Yuan Meng, and Peng Chang

College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China

Correspondence should be addressed to Xin-Rui Liu; liuxinrui@ise.neu.edu.cn

Received 14 December 2020; Revised 24 December 2020; Accepted 31 December 2020; Published 16 January 2021

Academic Editor: Rui Wang

The study of cyber-attacks, and in particular the spread of attack on the power cyber-physical system, has recently attracted considerable attention. Identifying and evaluating the important nodes under the cyber-attack propagation scenario are of great significance for improving the reliability and survivability of the power system. In this paper, we improve the closeness centrality algorithm and propose a compound centrality algorithm based on adaptive coefficient to evaluate the importance of single-layer network nodes. Moreover, we quantitatively calculated the decouple degree of cascading failures caused by exposed nodes formed by attack propagation. At last, experiments based on the IEEE 57 test system show that the proposed compound centrality algorithm can match the cyber-attack propagation scenario well, and we give the importance values of the nodes in a specific attack scenario.

1. Introduction

In recent years, electric power systems are facing unprecedented threat from cyber-attacks due to the rapid development of the information network and the higher integration of critical infrastructure and IED (intelligent electronic device) equipment in power CPS (cyber-physical system) [1–3]. The widely deployed ICT (information and communications technology) system makes the interaction more complex among multiple systems, especially more vulnerable under cyber-attacks. The failure of a single-layer system after being attacked by a cyber-attack will spread through the interdependent network, causing fragmentation and cascading failures [4, 5] and induce a large-scale power flow abnormal transferring. The abovementioned cascading failure process and cyber-attacks will eventually lead to a blackout that led to the collapse of the power system such as large blackouts which occurred in North America in 2003 [6], Rome in 2004 [7], and Ukraine in 2015 [8].

A large number of studies have shown that the power grid has small-world effect and scale-free property [9]. It shows strong robustness under random attacks but very fragile to deliberate attacks. Wang et al. in [10, 11] proposed methods to improve the stability of the power system. Albert et al. in [12] found that the power grid can maintain stable under most disturbances, but when the key power nodes are attacked, the synchronization ability of the grid will be greatly reduced. Therefore, identifying and assessing important nodes in the power grid and performing prevention and control are of great significance for improving the reliability and survivability of the power system. The current node importance evaluation methods include evaluation methods based on local information such as K-shell decomposition method [13], based on node path such as closeness centrality [14], based on feature vectors such as PageRank algorithm [15], and based on node removal [16] and contraction [17]. The authors in [17] applied the node contraction method to the power grid and verified the feasibility of identifying important nodes based on the topological structure of the power grid. A comprehensive evaluation index that takes into account both electrical characteristics and topological structure characteristics is proposed to identify important nodes in the power grid [18]. Based on the node link strength defined by power flow tracking, the author in [19] identifies important nodes in the power system from the perspective of global energy

transmission. In [20], the authors analyze the benefits, losses, costs, and other factors of network attacks and use dynamic Bayesian networks to comprehensively evaluate the attack effects of network nodes. In [21], the improved threat propagation tree is used to evaluate the situation and a CPPS security situation assessment model that considers threat propagation is proposed. The above node importance evaluation method only focuses on the characteristics of a single-layer network and does not consider the method of identifying key nodes of the power system under the interdependent network.

About interdependent networks, Buldyrev first analyzed the cascading failure in 2010 and proposed a cascading failure model based on network topology [22]. The author in [23] proposed electrical distance and node electrical coupling connectivity metric to identify key nodes in complex power grids. The average load balance of adjacent nodes and the parameter of network load rate are combined to measure the impact of the disabled node on the network and judge the importance of nodes in the interdependent network [24]. The above node importance evaluation considers the impact of disabled nodes on the interdependent network from different aspects but does not effectively analyze the composite value of nodes in the power CPS under the cyber-attack propagation scenario.

Whether the above node importance evaluation is based on a single-layer network or an interdependent network, the importance value is only a fixed value calculated based on a certain characteristic of the system. Without considering the potential dangers brought by the spread of cyber-attacks, such fixed indexes cannot meet the system's need to distinguish important nodes under the attack spread scenarios. In this paper, the analysis of potential dangers caused by the spread of cyber-attacks is shown in Figure 1. Suppose that at a certain time during the interval between two detections, a certain device in the information network is attacked by a cyber-attack (such as worms, and Trojan Horse).

(1) First, before the attacked device is successfully detected by the power CPS and the countermeasures are executed, the spread of the cyber-attack is in the first stage. In this stage, the attack may have spread from the initially attacked device to the remaining devices that have topological connections or information interactions with it. Take the information layer network as an example: the data detection frequency cannot match the information exchange rate, which will cause the cyber-attack to be spread to the new information equipment due to the information exchange before the next detection. If the attack makes the transmitted device in exposed state (the device has been successfully attacked, but the attacker did not perform any attack operations), then the power system cannot successfully detect such exposed device in the next detection.

(2) Subsequently, the power system detects the initial attacked device and executes countermeasures. Although the current detection failure rate can meet the expectations of the dispatch center, it cannot be reduced to zero. Therefore, the response strategy made by the dispatch center can only reduce and block the harm caused by cyber-attacks to a certain extent. And, this strategy does not consider the cyber-attack propagation described in (1). In fact, after the first propagation stage, there are most likely exposed nodes in the power system.

(3) The period from the execution of the strategy to the next detection, the spread of cyber-attacks has reached the second stage. Although the previous strategies dealt with the initially attacked device, the exposed nodes formed in the first propagation stage will continue deepen the scope of cyber-attacks over time.

(4) After the above two propagation stages, most devices have the potential to be attacked. This potential risk may expose the power system to great danger during the interval between two detections.

In order to improve the proposed node importance evaluation method of power CPS and effectively reduce the potential risks caused by the cyber-attack propagation between the detection gap, in this paper, from the information network and cyber-physical interdependent network two aspects, we propose a node importance evaluation method suitable for cyber-attack scenarios. The main contributions of this paper are as follows:

(1) At the information network level, considering the propagation of cyber-attacks among information devices, a compound centrality index and its calculation method of nodes based on adaptive coefficients are proposed. This index improves the one-sidedness of the existing centrality indexes, and the time-varying calculation result better matches the importance of the power CPS node under the cyber-attack propagation scenario.

(2) Establish a cyber-physical interdependent network model to analyze the cascading failure behavior characteristics of the system and give a quantitative analysis of the degree of power grid decoupling under different attack scenarios. Calculate the degree of decoupling caused by cascading failures which caused by exposed nodes under the spread of cyber-attacks and comprehensively reflect the importance of different nodes in power CPS.

In the second section, we proposed the node risk and established a network attack propagation model. The third section improves closeness centrality and proposes a compound centrality algorithm based on adaptive coefficients. In the fourth section, the power system cyber-physical interdependent network is established to quantitatively calculate the degree of cascading failure that the exposed node formed by the spread of cyber-attacks may cause. In the fifth section, a case study is given to prove the advantages of our proposed algorithm and report the conclusion.

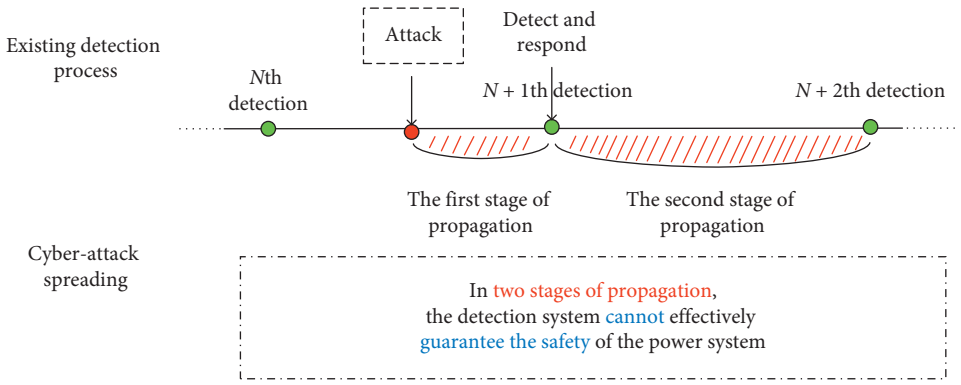


Figure 1: Potential risks under the spread of cyber-attacks.

2. Cyber-Attack Propagation Model

Cyber-attacks occur more commonly in power communication networks due to the deep integration of communication and measurement devices such as phasor measurement units (PMUs). Attackers use new viruses and worms with autonomous propagation capabilities to falsify or control the resources of the information layer to endanger the security and reliability of the power grid. Therefore, in this section, establish a cyber-attack propagation model for power CPS information layer devices which can apply to all equipped information devices with communication capabilities, such as information acquisition system SCADA (Supervisory Control and Data Acquisition), energy management system EMS (Energy Management System), WAMS (Wide Area Measurement System), and so on.

Since SCADA, EMS, and WAMS have a certain time interval in collecting, monitoring, processing information, and transmitting instructions, while considering the characteristics of information systems based on discrete events and the propagation mechanism of new viruses such as worms, the unit propagation time of the attack $\hat{t}$ is defined as the data interaction interval of power CPS. According to the propagation stage of the cyber-attack, the information equipment is divided into 5 states. A single device can only be in one state within a unit of propagation time $\hat{t}$. Under certain conditions, the current state can be transformed into other states. The situation transformation process is shown in Figure 2.

(1) *Infected*. The device has been successfully infected by a cyber-attack and has been manipulated by the attacker to perform some actions (for example, tampering with operating data and modifying switch status). It has the ability to infect other devices. The detection system can detect the device.

(2) *Exposed*. The device successfully infected by a cyber-attack, but the attacker has not performed any attack operations. It has the ability to infect other devices. The detection system cannot find the device.

(3) *Susceptible*. The device directly connected to the infected or exposed device in the communication topology. There are security vulnerabilities that can be infected by malicious code such as worms but have not been infected yet and do not have the ability to infect other devices.

(4) *Isolated*. The device that has been attacked is automatically isolated by the detection software or manually by the system operator. No longer have hardware connections or data interactions with other devices. No longer has the ability to infect other devices.

(5) *Normal*. Devices are in a normal state of untouched cyber-attacks.

2.1. Probability of Propagation. In this section, we calculate the cyber-attack propagation probability between pairs of individuals who have a connection in the communication topology. Consider a pair of devices which are connected, one of which i is infective or exposed and the other j susceptible. Attack spreads through them by contact from i to j. Then, the rate that j not being attacked in the continuous time system is as follows [25]:

$$1 - \gamma_{ij} = \lim_{\delta t \longrightarrow 0} \left(1 - r_{ij}\delta t\right)^{\tau_i/\delta t} = e^{-r_{ij}\tau_i}, \tag{1}$$

where r_{ij} is the probability that the cyber-attack successfully spreads from infective or exposed device-i to susceptible device j under the self-protection mechanism of the device and information system. The infective device i remains infective for a time τ_i that i has not been detected and disabled by the detection system.

Due to the cyber systems based on discrete events, use discrete time-steps rather than continuous time, in which case instead of taking the limit in equation (1) we simply set $\delta t = 1$, giving

$$\gamma_{ij} = 1 - \left(1 - r_{ij}\right)^{\tau_i}, \tag{2}$$

where τ_i is the average detection time of the online detection system that is measured in data transmission time-steps.

2.2. Propagation Model. For modeling the attack propagation clearly, we abstract different types of communication devices in the information layer as nodes. According to the

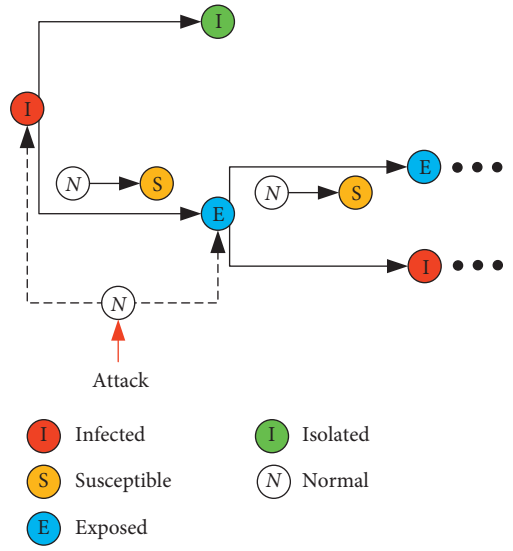


Figure 2: The devices state transition under cyber-attack.

communication topology of the information network, the connection matrix $C = c_{ij}$ is defined. The matrix elements c_{ij} are binary variables which equals 1 if i is kept connected to j in communication and 0 otherwise. We define the probability that a node may be attacked and turn into exposed state as the node risk $P(\cdot)$. Let us assume at the initial time t_0 between two detection gaps, one or several nodes in the CPS information layer are compromised by cyber-attacks and turn into infected while the remaining nodes are uncompromised. In the following period of time, the remaining nodes will turn into exposed with a certain probability, accompanied by a time-varying node risk.

We define $\psi = \{1, 2, \ldots, M\}$ as the set of information nodes; $\eta = \{\ldots\}$ is the set of infected nodes; and $T = n\widehat{t}$ is the time when the attack spreads through n units of propagation time $\widehat{t}$. The attack spreads in stages, and the risk of each node changes with time as follows [26]:

(1) At the initial time t_0, the risk of infected and other nodes are as follows:

$$\begin{aligned} P\left(\alpha_i^{t_0} = 1\right) &= 1, \quad \forall i \in \eta, \\ P\left(\alpha_i^{t_0} = 1\right) &= 0, \quad \forall i \notin \eta, \end{aligned} \tag{3}$$

where α_i^t is a binary variable which equals 1 ($\alpha_i^t = 1$) if the node i has the possibility of being attacked at time t_0, and $\alpha_i^t = 0$; otherwise, $P(\alpha_i^{t_0} = 1)$ is the node risk of i at time t_0.

(2) By time $t_0 + \widehat{t}$, cyber-attacks may spread from infected nodes to susceptible nodes via routers and communication connections, turning susceptible nodes into exposed nodes, and the detection software has not detected them yet. The risk of each node is given by the following equation:

$$P\left(\alpha_i^{t_0+\widehat{t}} = 1\right) = 1, \quad \forall i \in \eta, \tag{4}$$

$$P\left(\alpha_i^{t_0+\widehat{t}} = 1\right) = 1 - \prod_{j\in\eta}\left(1 - \gamma_{ji}\right), \quad \forall i \notin \eta, \tag{5}$$

where γ_{ji} is the probability that the attack propagates from node i to node j during the time $\widehat{t}$, and it is given in equation (2).

(3) By time $t_0 + 2\widehat{t}$, the range and volume of cyber-attacks further expand over time, and the risk of exposed nodes deepens. The risk of each node is an iterative process over time. According to equations (4) and (5), the node risk by time $t_0 + 2\widehat{t}$ is given by the following equation:

$$P\left(\alpha_i^{t_0+2\widehat{t}} = 1\right) = 1, \quad \forall i \in \eta, \tag{6}$$

$$\begin{aligned} P\left(\alpha_i^{t_0+2\widehat{t}} = 1\right) = {} & P\left(\alpha_i^{t_0+2\widehat{t}} = 1 | \alpha_i^{t_0+\widehat{t}} = 1\right) \times P\left(\alpha_i^{t_0+\widehat{t}} = 1\right) \\ & + P\left(\alpha_i^{t_0+2\widehat{t}} = 1 | \alpha_i^{t_0+\widehat{t}} = 0\right) \times P\left(\alpha_i^{t_0+\widehat{t}} = 0\right), \quad \forall i \notin \eta, \end{aligned} \tag{7}$$

where $P(\alpha_i^{t_0+2\widehat{t}} = 1|\alpha_i^{t_0+\widehat{t}} = 0)$ is the probability that the node is not attacked at time $t_0 + 2\widehat{t}$ but is attacked at time $t_0 + \widehat{t}$. In equation (7), only $P(\alpha_i^{t_0+2t} = 1|\alpha_i^{t_0+t} = 0)$ is unknown during the iteration process. This variable is discussed as follows:

$$\begin{aligned} &P\left(\alpha_i^{t_0+2\widehat{t}} = 1|\alpha_i^{t_0+\widehat{t}} = 0\right) \\ &= 1 - P\left(\alpha_i^{t_0+2\widehat{t}} = 0|\alpha_i^{t_0+\widehat{t}} = 0\right) \\ &= 1 - \prod_{\substack{i,j\notin\eta \\ i\neq j}} \left(1 - P\left(\alpha_j^{t_0+\widehat{t}} = 1\right) \times \gamma_{ji}\right), \quad \forall i, j \in \psi. \end{aligned} \tag{8}$$

Substituting equation (8) into equation (7) to obtain the risk of some nodes at time $t_0 + 2\widehat{t}$:

$$\begin{aligned} P\left(\alpha_i^{t_0+2\widehat{t}} = 1\right) &= 1 \times P\left(\alpha_i^{t_0+\widehat{t}} = 1\right) \\ &+ \left(1 - \prod_{\substack{i,j\notin\eta \\ i\neq j}} \left(1 - P\left(\alpha_j^{t_0+\widehat{t}} = 1\right) \times \gamma_{ji}\right)\right) \\ &\times \left(1 - P\left(\alpha_i^{t_0+\widehat{t}} = 1\right)\right), \quad \forall i, j \in \psi, i \notin \eta. \end{aligned} \tag{9}$$

(4) By time $t_0 + n\widehat{t}$ $(n > 2)$, the node risk is given as follows:

$$P\left(\alpha_i^{t_0+n\widehat{t}} = 1\right) = P\left(\alpha_i^{t_0+n\widehat{t}-1} = 1\right) + \left(1 - \prod_{\substack{i,j\notin\eta \\ i\neq j}} \left(1 - P\left(\alpha_j^{t_0+n\widehat{t}-1} = 1\right) \times \gamma_{ji}\right)\right) \times \left(1 - P\left(\alpha_i^{t_0+n\widehat{t}-1} = 1\right)\right), \quad \forall i, j \in \psi. \tag{10}$$

(5) During the above propagation process, if the initially infected node is detected and processed by the power system at time $t = t_0 + x\widehat{t}$, then in the time after that, the risk of the node is

$$P\left(\alpha_i^{t_0+(x+\cdots)\widehat{t}} = 1\right) = 0, \quad \forall i \in \eta. \tag{11}$$

Notice. One of the ways the power system responds to cyber-attacks is as follows. After successfully detecting the cyber-attack, the power system will disable infected nodes automatically by the detection software or manually by the system operator. These nodes are no longer connected to the information network (no more data interaction with other nodes), while the system no longer trusts its uploaded data. Namely, this type of node turns into the isolated state. Then, they may be recovered to normal by the system and reconnect to the network, but in the propagation discussed in this paper, such nodes no longer participate in the propagation of cyber-attacks, that is, they cannot be reinfected by cyber-attacks.

3. Node Centrality Algorithm Based on Adaptive Coefficient

The connection form of the node and its position in the topological structure have a vital influence on the promotion or interruption of the spread of cyber-attacks [4]. At present, a variety of centrality algorithms have been applied to node importance evaluation. In this section, we propose a node importance evaluation algorithm that considers cyber-attacks propagation and the potential threats it may cause in the detection gap, which is more flexible and more suitable for attack scenarios.

3.1. Centrality Analysis. According to the connection form of the node and its position in the topology, common node centrality algorithms include degree centrality [27], closeness centrality [28], betweenness centrality [28], and so on, all describe the importance of nodes in the network from different respects, and they are given as follows:

$$\mathrm{DC}(i) = \sum_{j=1, j\neq i}^{N} c_{ij}, \tag{12}$$

$$\mathrm{CC}(i) = \frac{N-1}{\sum_{j=1, j\neq i}^{N} \mathrm{dis}_{ij}}, \tag{13}$$

$$\mathrm{BC}(i) = \sum_{\substack{s\neq t\neq i\in V \\ s<t}} \frac{\sigma_{st}(i)}{\sigma_{st}}, \tag{14}$$

where N is the number of nodes in the network, c_{ij} is a binary decision variable that judges the topological structure or information interaction between nodes i and j, dis_{ij} is the shortest path between nodes i and j; in this paper, we use the minimum number of nodes in the path from i to j as the value of the shortest path.σ_{st} is the number of shortest paths from s to t; $\sigma_{st}(i)$ is the number of shortest paths from s to t which passing by node i.

Degree centrality indicates the sum of the number of nodes directly connected to the designated node. Closeness centrality is the reciprocal of the average shortest path from the designated node to all other reachable nodes. In general,

the closer a node is to other nodes, the greater its closeness centrality. Conversely, the smaller the node closeness centrality, the more the node is at the edge of the network. Betweenness centrality is the number of times that the designated node is located on the shortest path between any two other nodes.

The three algorithms describe the importance of nodes from the perspectives of local characteristics, global characteristics, and propagation characteristics, while they also have certain limitations. Degree centrality can only one-sidedly reflect the closeness between the designated node and its surrounding nodes. For example, the connection node between two partitioned networks has a smaller degree of centrality but a higher degree of importance. Or for some nodes with high degree centrality, the clustering network it embeds may be at the edge of the system. At this time, the degree centrality is relatively high but the degree of importance is average. However, as the nodes in a complex network, especially in large-node systems such as power systems, the position of the node in the system and some of its functions are far more important than the number of nodes around it. The descending order curve of the normalized values of DC, CC, and BC of each node in the IEEE14 system is shown in Figure 3. The comparative analysis is as follows:

(1) The relative DC values of all nodes are arranged in descending order as shown in. It descends in steps, and there are 3 steps and some steps are larger in width. It descends in steps, and there are 3 steps and some steps are larger in width. For example, the DC of 6 nodes in the graph are all 0.4. The importance of these nodes in the same step cannot be distinguished.

(2) The descending arrangement curve of relative CC is relatively smooth, but there is little difference among the values. The variable interval of the value is 0.585–1, which is too narrow.

(3) The relative BC descending order curve is also a smoother curve with values distributed in the interval 0-1. However, there is a platform segment with a value of 0, which accounts for 28.57% of the total system nodes, at its end. The node in this segment cannot effectively distinguish the importance.

3.2. Centrality Algorithm Based on Adaptive Coefficient. According to the analysis results of the three centrality algorithms in 3.1, we can get the following. The three algorithms have their respective advantages and disadvantages. If only one centrality algorithm is selected as the basis for judging the importance of nodes, the sorting result is not appropriate and accurate. These centrality algorithms make it impossible to effectively distinguish the importance of each node in the system under a specific attack scenario based on the fixed value calculated by the topology structure. Moreover, if the importance of nodes is only distinguished based on the node risk P that proposed in 2.2, the position of the node in the system is ignored, and the mutual influence between nodes in the process of attack propagation is discarded. Therefore, after comprehensively considering the global attributes of CC and the propagation attributes of BC, this section improves the CC algorithm according to the propagation characteristics of cyber-attacks and proposes a compound centrality algorithm that considers both improved CC and BC to weaken the differences and shortcomings of existing centrality algorithms.

Improved closeness centrality algorithm:

$$\mathrm{CC}(i)' = \sum_{j\in\theta, i\in\psi-\theta} \mathrm{dis}_{ji}, \tag{15}$$

where dis_{ji} is the shortest path from j to i, j is the node initially infected by the cyber-attack, and i is the remaining nodes in the system.

The compound centrality algorithm proposed in this paper performs a weighted summation of the two indexes of improved CC and BC, but the two are not completely independent indexes. For example, nodes with a high improved CC will have a higher BC under a certain probability. Summing the two indexes with a weight of 0.5, reasons such as duplication of information and redundant factors will cause the calculated composite index to be inaccurate. To solve this problem, this section proposes an adaptive coefficient as the weight of the two centralities.

In the propagation of the power CPS cyber-attack, the propagation probability between nodes γ_{ij}, as a key factor, has always been defined as a fixed value in previous studies. Considering the differences in the importance of each node in the network, a correction equation is proposed to modify γ_{ij} to a certain extent:

$$\gamma_{ij}' = \gamma_{ij} - X * C_j, \tag{16}$$

where X is the correction coefficient and C_j is the centrality index of node j. The γ_{ij}' of node j changes with the centrality index C_j. The higher the C_j, the lower the γ_{ij}'. That is, when C_j increases, the probability of a cyber-attack spreading from node i to node j decreases.

The iterative solution process of the adaptive coefficient ∂ is given by equations (17)–(23). Equation (22) is the termination condition of the iteration, and equation (23) is the objective function. The appropriate initial values of ∂_1 and ∂_2 are selected and substituted into formula (17) to start the iteration. In equations (18) and (19), the composite centrality index C of the node and the propagation probability γ_{ij} are updated with the update of ∂. According to the risk model proposed in 2.2, use the iteratively updated γ_{ij}' to calculate the risk of each node and the average of the overall node risk, which is shown in equation (21). Substituting the average of the overall node risk in the current and previous iterations into equation (17) again will start a new round of iterative correction process. Until ∂ satisfies the iteration termination condition $\partial_k - \partial_{k-1} \le 10^{-5}$, the iteration process stops and jumps out of the iteration. Take the minimum the average of the overall node risk in all rounds of iterations as the objective function and find the optimal solution that satisfies the objective function in the iteration process. That is, at the target time $t = T$ (T is an integer multiple of $\hat{t}$), after correcting with the optimal ∂, the average of the overall node

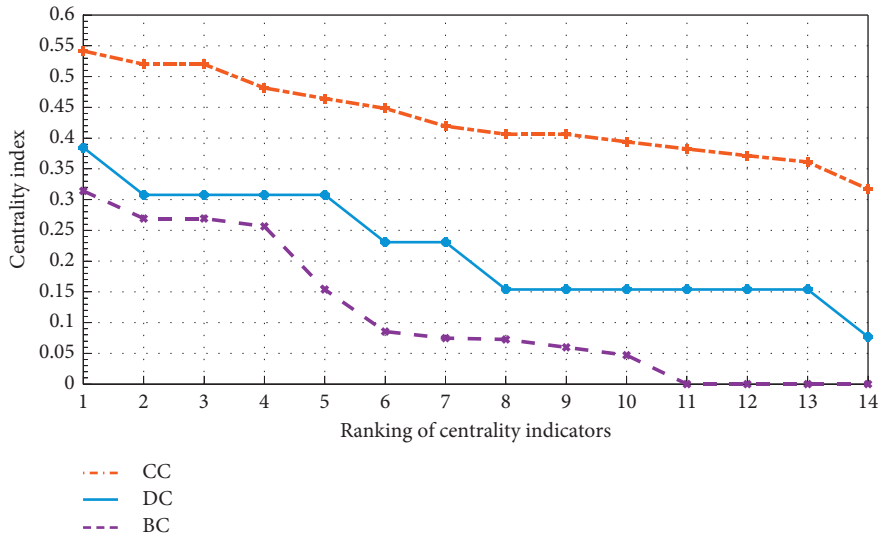


Figure 3: Central index descending order of the IEEE14 system.

risk in the system is the smallest. This optimal solution is the adaptive coefficient of this paper. It can be seen from the above algorithm flow that if we want to start the iterative process of ∂, we need to give the first two initial values ∂_1 and ∂_2 of ∂. The rationality of the initial value selection and its influence on the number of iterations and the judgment of the optimal solution are discussed in the case study in Section 5.

$$\partial_k = 1 + \frac{\left|\overline{P_{k-1}^T} - t\overline{P_{k-1}^{T-1}}\right|}{\left|\overline{P_{k-2}^T} - t\overline{P_{k-2}^{T-1}}\right|}, \tag{17}$$

$$C_i^k = \frac{1}{\partial_k} bc_i + \left(1 - \frac{1}{\partial_k}\right) cc_i', \tag{18}$$

$$\gamma_{ij}^k = \gamma_{ij} - X * C_j^k, \tag{19}$$

$$P\left(\alpha_i^{t_0+n\widehat{t}} = 1\right) = P\left(\alpha_i^{t_0+n\widehat{t}-1} = 1\right) + \left(1 - \prod_{\substack{i,j \notin \eta \\ i \neq j}} \left(1 - P\left(\alpha_j^{t_0+n\widehat{t}-1} = 1\right) \times \gamma_{ji}\right)\right) \times \left(1 - P\left(\alpha_i^{t_0+n\widehat{t}-1} = 1\right)\right), \quad \forall i, j \in \psi, \tag{20}$$

$$\overline{P_k^T} = \frac{\sum_i P_i^T|k}{N}, \tag{21}$$

$$\partial_k - \partial_{k-1} \leq 10^{-5}, \tag{22}$$

$$f = \min_k \overline{P_k^T}, \tag{23}$$

where ∂_k is the adaptive coefficient of the kth iteration, C_i^k is the compound centrality of node i when $\partial = \partial_k$, γ_{ij}^k is the probability of successful propagation of the attack from i to j modified according to C_i^k, $P_i^T|k$ is when $\gamma_{ij} = \gamma_{ij}^k$, the risk of node i at time T, $\overline{P_k^T}$ is when $\gamma_{ij} = \gamma_{ij}^k$, the average risk of all nodes in the system at time T, and f is the decision function of ∂. A certain ∂_k in the iterative process minimizes the average risk $P_i^T|k$ of nodes in the system at the target time T.

4. Cascading Failure of Power Interdependent Network

4.1. Interdependent Network Model. Electric power CPS is a multidimensional heterogeneous system that deeply embeds perception, information processing, and control platforms into the power system to meet real-time monitoring and achieve command-driven of the power system. The power network provides power for the information network, and the control and analysis module in information network reversely drives the power network. The interdependence between the two networks enables the power system to be modeled as a power cyber-physical interdependent network. Based on the "undirected" and "disordered" characteristics of general types of cyber-attacks (nondirected attacks) spreading in information systems, both the power network and the information network are equivalent to nonweighted undirected networks. Combining the actual situation of China's power system [5], only considering the characteristics of interconnection between nodes, using complex network theory to simplify the power and information network is as follows:

(1) Ignore the functional differences among the plants and stations, and regard the power generation nodes and substation nodes, dispatching nodes and routing nodes in the information network as equivalent nodes, regardless the difference in the types,

quantities, and deployment modes of devices in various sites

(2) Ignoring the differences in information protocols and hierarchical structures among information nodes at all levels, it is considered that the lines between nodes can transmit bidirectional information, and multiple information lines in the same direction are combined to eliminate multiple edges and self-loops.

Use $G_P = (E_P, V_P)$ and $G_C = (E_C, V_C)$ to represent a power network with n nodes and k branches and an information network with m nodes and g branches, where $E = \{e_{ij}\}$ is the set of edges and $v = \{v_1, v_2, \ldots, v_n\}$ is the set of nodes in the network, respectively. Each node in the power network G_P is connected to an information node in G_C. The power node provides electrical energy support to the information node, and the information node receives the status information sent by the power node and feeds back control instructions. Naturally, each node in G_P connects and depends on the corresponding node in G_C, and vice versa. In addition, according to the important status of the scheduling nodes, it is set as an autonomous node independent of the power grid. It deploys a complete backup power supply and power generation equipment, which is not affected by power grid energy fluctuations.

China's power line information network is a dedicated resource for the power system, and most of the communication lines are laid along with high-voltage transmission lines. The geographical similarity of the two layouts makes the topological structure between the information network and the power network highly similar. On the other hand, in order to meet the needs of control and dispatch, the information network also has dispatching nodes. So, the information network has more stations than the power network. At the same time, the optical information network needs to be formed into a ring to protect its self-healing ability, and the structure of the information network is more complicated [29]. Therefore, the dependent network model selects the "part-to-one correspondence" coupling mode, and the established power cyber-physical dependent network model is shown in Figure 4.

Different from the single-layer network with only connectivity link, there are two types of edges in the interdependent network: the connectivity link and the dependency link. The nodes in the single-layer network rely on the internal connectivity link (the black solid lines in Figure 4) to achieve corresponding functions. For example, the power generation nodes, substation nodes, and load nodes in the power network realize the generation, transmission, and consumption of electric energy through the transmission line. The topological structure of the information network is more complicated than that of the power network. In Figure 5, there are connecting links that do not exist in the power network between the information node 2 and the information node 8. The dependency link between the two networks (the red dotted line in Figure 4) is used as a medium for energy or information exchange to realize the mutual influence between the two-layer networks. The dependency matrix D_{P-C}, D_{C-P} is defined according to the connection relationship of the power cyber-physical dependent network, and the matrix element d_{ij} uses logic elements of "1" and "0" to indicate whether there is a dependency edge between the power and information nodes.

In summary, establish a power cyber-physical interdependence network that includes power grids, information networks, and the interdependency links $G_P = G_C = \Re(G_P, G_C, D_{C-P}, D_{P-C})$, where $G = (E, V)$,$V = \{v_1, v_2, \ldots, v_n\}$ is the set of nodes in a single-layer network, $E = \{e_{ij}\}$ is the set of connectivity links in a single-layer network, and D_{P-C}, D_{C-P} is the set of the dependency links between the power cyber-physical dependent network.

4.2. Cascading Failure Behavior. According to the operating principle of the dependent network, the failure of the power node or information node will lead to the failure of the dependent node in the other network. The power network or information network will be broken into several fragmented networks, and the scope of the attack will expand with the expansion of the network fragmentation and eventually cause cascading failures in the dependent network [29]. Describe the cascading failure behavior caused by the attack as follows:

(1) When one or some information nodes (power nodes) in the information network G_C (or power network G_P) are attacked and fail, the connectivity links and the dependency links on these nodes fail too;

(2) Corresponding nodes in the power network G_P fail due to the interdependence with the failure node in G_C. Corresponding, the connectivity links and the dependency links on the nodes also fail. After removing all the abovementioned faulty nodes, the connectivity links, and the dependency links, the power and information network is decomposed into several fragmented networks.

(3) (On the basis of 2), identify the connected subgraph in G_P and G_C, and judge the nodes that do not belong to the connected subgraph as failed nodes. Remove the newly determined failed node and its connectivity links and dependency links. Based on this process, the system will reach stability after a certain number of propagations of the failure in the dependent network.

(4) Identify the set of nodes in the maximal connected subgraph in the stable network, and finally determine the stable state of the decoupled network.

Notice. According to the main research content of this paper, the information layer network attack and its propagation mechanism demand for network connectivity use "the set of nodes in the maximal connected subgraph" [29] to judge the stable state of the dependent network. The set element is the power-cyber node group $v_P - v_C$, and the dependency links between v_P and v_C, where v_P belongs to the maximal

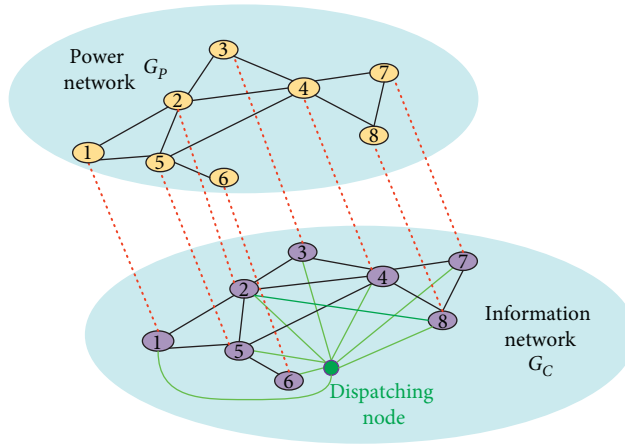


Figure 4: Example of the power CPS interdependent network.

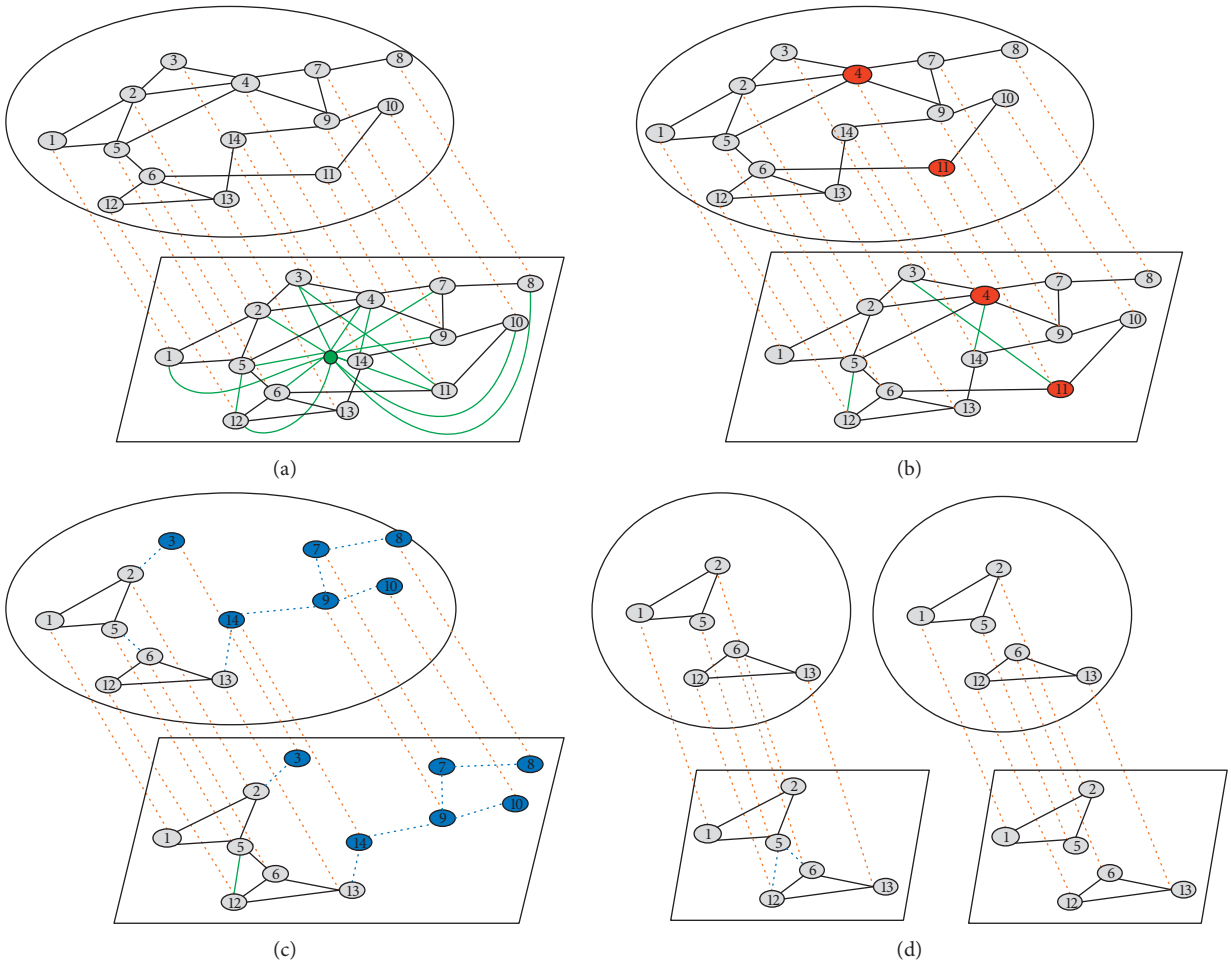


Figure 5: Cascading failure behaviors of the power interdependent network.

connected subgraph in the power network and v_C also belongs to the maximal connected subgraph in the information network. In order to ensure the effectiveness of connectivity, after the fragmentation of the network, only the nodes in the set of nodes in the maximal connected subgraph are available, and the remaining nodes fail.

According to the cascading failure behavior, taking the IEEE14 node as an example, the failure decoupling process of the dependent network is shown in Figure 6 when the initial attack node is 4 and 11. The upper network in the figure is the power system G_P, and the lower network is the information system G_C. Figure 6(a) shows that the

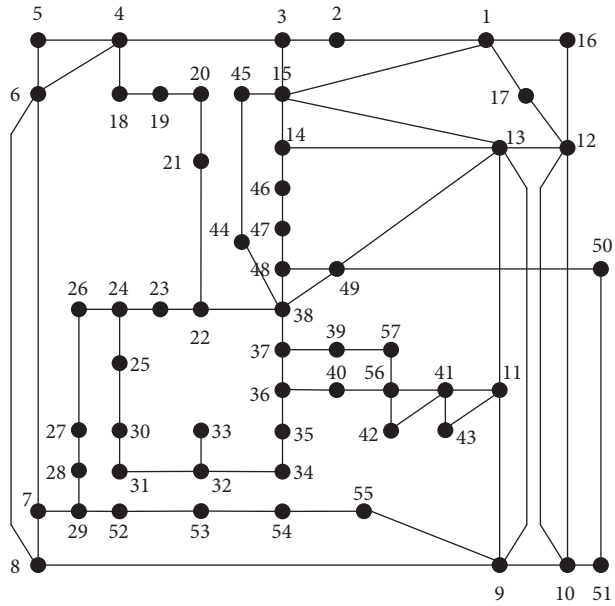


Figure 6: The IEEE57-bus test system.

connectivity of G_C is similar to that of G_P, but compared with G_P, G_C has one more independent green "scheduling node" and green connectivity links between the scheduling nodes and each information node, and three more green connectivity links between 4–14, 3–11, and 5–12. In Figure 6(b), the nodes 4 and 11 failed due to cyber-attacks and turned to red. Correspondingly, the connectivity link connected to it fails and turns to a dashed line, and the dependency link fails and turns to a red line, and the corresponding dependent nodes and connectivity links in G_P are treated in the same way. In Figure 6(c), we delete the failed node and the failed edge in Figure 6(b). The remaining nodes that do not belong to the connected subgraph are turned into blue, and the connecting edges on these nodes turn to blue dashed lines. At this time, it can be seen that both networks are fragmented since 5–12 connecting edge in G_C, and the number of nodes in the connected subgraph is greater than that in G_P, and the degree of fragmentation is not as serious as G_P. Figure 7(d) filters out the set of nodes in the maximal connected subgraph from Figure 7(c). At this time, the power interdependent network finally reaches a stable state.

4.3. Decoupling Degree Based on Cascading Failure. According to the behavior of cascading failures, it can be seen that the cascading failure propagation process in this paper is suitable for random cyber-attacks and deliberate cyber-attacks of power CPS. The attacker can choose one or some nodes as the attack target to form different combinations of cyber-attacks. Different combinations of attacks will cause different decoupling processes. The exposed nodes formed in the process of cyber-attack propagation will not only cause the attack propagation between the information layer devices during the detection gap but also cause the cascading failure of the power cyber-physical dependent network. The node survival rate S is established to describe the degree of decoupling of power CPS when the system reaches a steady state after the cascading failure. The larger the value of S, the more the number of nodes remaining in the stable state of the system and the lower the degree of system decoupling.

$$S = \frac{N_C' + N_P'}{N_C + N_P}, \tag{24}$$

where N_P and N_P' are the number of effective nodes in the power network before and after the cascade failure and N_C and N_C' are the number of effective nodes in the information network before and after the cascade failure.

5. Experimental Results

We use the connection relationship of the IEEE57-bus test system to simulate the information layer network connection relationship, ignoring the weight of each link; it is shown in Figure 5.

5.1. Compound Centrality Algorithm

Notice. In the information layer cyber-attack propagation model described in Section 1, the initial time t_0 should be the time when the cyber-attack occurred. However, at present, the power system cannot accurately analyze which time the cyber-attack occurred between the two detection gaps. Therefore, in the experiment, the moment when the cyber-attack is detected is regarded as the initial moment in the risk $P(\cdot)$ calculation. And, set the power system to make a response strategy to disable the attacked device at time $t + \hat{t}$.

Establish the attack scenario with attack the information device equipped on node 18. The BC and improved CC

Table 1: Values of BC and improved CC under IEEE57 attacking the 18th node.

Importance ranking	Node	Improved CC	Node	BC	Importance ranking	Node	Improved CC	Node	BC
1	4	1	38	493.47	29	49	5	21	82.917
2	19	1	13	432.47	30	52	5	30	68.245
3	3	2	9	373.28	31	55	5	40	67.543
4	5	2	49	337.51	32	24	6	44	58.817
5	6	2	22	294.18	33	27	6	48	55.767
6	20	2	37	255.63	34	37	6	20	55.5
7	2	3	11	221.13	35	41	6	45	55.15
8	7	3	36	218.67	36	43	6	54	51.742
9	8	3	24	209.83	37	47	6	14	50.233
10	15	3	15	204.59	38	48	6	31	49.317
11	21	3	8	189.39	39	50	6	19	48.083
12	1	4	23	177.93	40	51	6	1	43.183
13	9	4	41	161.46	41	53	6	10	36
14	13	4	29	157.58	42	54	6	39	31.75
15	14	4	7	153.08	43	25	7	50	30.602
16	22	4	35	149.09	44	26	7	52	29.7
17	29	4	4	134.47	45	36	7	53	26.442
18	45	4	56	124.46	46	39	7	47	21.75
19	10	5	3	112.92	47	42	7	57	20.25
20	11	5	34	112.59	48	56	7	46	18.983
21	12	5	6	111.95	49	30	8	51	13.602
22	16	5	12	107.50	50	35	8	2	9.758
23	17	5	25	104.49	51	40	8	16	6.667
24	23	5	32	99.155	52	57	8	17	6.667
25	28	5	28	99.15	53	31	9	5	0
26	38	5	26	94.733	54	34	9	33	0
27	44	5	55	92.242	55	32	10	42	0
28	46	5	27	88.483	56	33	11	43	0

Notice. According to the improved CC, the lower the CC, the closer the node is to the attacked node and the higher the node importance.

indexes of the remaining nodes calculated according to formulas (13) and (14) are shown in Table 1, and the two indexes are ranked according to importance. It can be seen from the table:

(1) The improved CC index mainly serves the attack scenario. The closer the node to the initial infected node, the lower the node's improved CC value. At the same time, the indexes have a certain degree of repeatability; for example, the value of the improved CC index for 7 nodes is equal to 4.

(2) The importance ranking results of BC and improved CC are obviously different. Some nodes have higher improved CC and lower BC, such as node 9. Some nodes have lower improved CC and higher BC, such as node 37.

Perform log normalization processing ($\widehat{x} = \log_{10}(x)/\log_{10}(\max)$) on the BC and improved CC indexes in Table 1, and use them as the input data to execute the compound centrality algorithm proposed in this paper. The iterative results ∂ and iteration times N of the adaptive coefficients under different correction coefficients X and different target times T are shown in Table 2.

According to the iterative results in Table 2, it can be seen that in the initial stage of attack propagation, there is a certain proportion of improved CC in the compound centrality. That is, in the initial stage of propagation, mainly modifying γ_{ij} of node j which is near the infected node has a better effect on reducing the overall risk. However, as the attack spreads, the proportion of BC in the compound centrality gradually increases. That is, the propagation characteristics possessed by nodes in the topological structure in the later stage of propagation have a more obvious effect on reducing risk which caused by the spread of the attack. In addition, it can be seen that the iteration times of this algorithm are less than 10, and the convergence speed of the algorithm is fast, which can meet the demand for timeliness of power system scheduling. And, we verified that the selection of the initial value of ∂ has no effect on the convergence value of the iteration and the optimal solution. It has a slight impact on the iteration times, but the number of iterations can also be guaranteed within 10 times.

In the case of $X = 0.03$, $T = 15\widehat{t}$ and $X = 0.03$, $T = 30\widehat{t}$, using compound centrality, improved CC, BC, and DC to modify γ, the risk of the remaining nodes of the system is shown in Figures 7 and 8.

Figure 7 shows the following:

(1) In the initial stage of attack propagation, the risk of several nodes that are topologically close to the initial infection node 18 rises fastest (for example, node 4, 19, 5, 6, 20)

(2) In the initial stage of attack propagation, in view of the strong correlation between the improved CC and

Table 2: Iteration results of the adaptive coefficients under the IEEE57 attacking the 18th node.

∂	Iteration times N						
	$T = 12\hat{t}$	$T = 15\hat{t}$	$T = 18\hat{t}$	$T = 21\hat{t}$	$T = 24\hat{t}$	$T = 27\hat{t}$	$T = 30\hat{t}$
$X = 0.03$	$\partial = 1.479$ $N = 6$	$\partial = 1.433$ $N = 7$	$\partial = 1.389$ $N = 6$	$\partial = 1.346$ $N = 8$	$\partial = 1.306$ $N = 6$	$\partial = 1.268$ $N = 9$	$\partial = 1.234$ $N = 7$
$X = 0.04$	$\partial = 1.429$ $N = 7$	$\partial = 1.385$ $N = 7$	$\partial = 1.342$ $N = 7$	$\partial = 1.301$ $N = 7$	$\partial = 1.263$ $N = 8$	$\partial = 1.227$ $N = 7$	$\partial = 1.194$ $N = 6$
$X = 0.05$	$\partial = 1.330$ $N = 6$	$\partial = 1.292$ $N = 6$	$\partial = 1.256$ $N = 6$	$\partial = 1.221$ $N = 7$	$\partial = 1.189$ $N = 6$	$\partial = 1.159$ $N = 7$	$\partial = 1.132$ $N = 6$

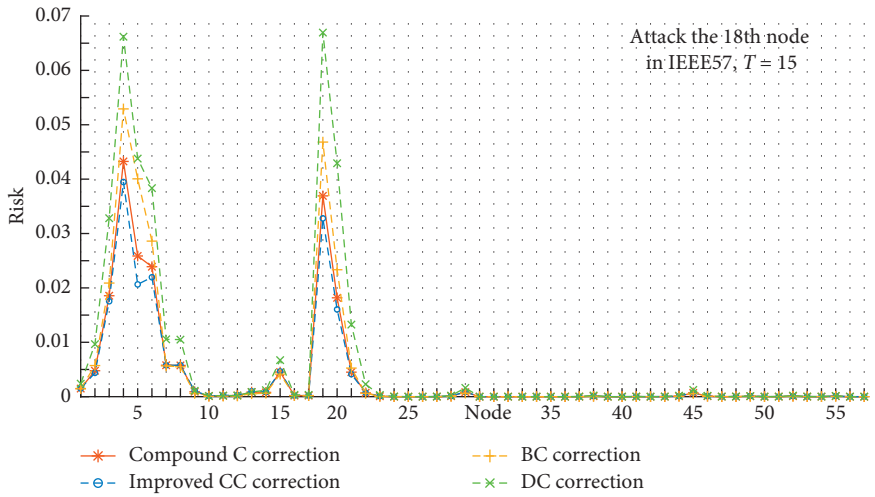


Figure 7: Risk of node under compound C, improved CC, BC, and DC correction at $T = 15$.

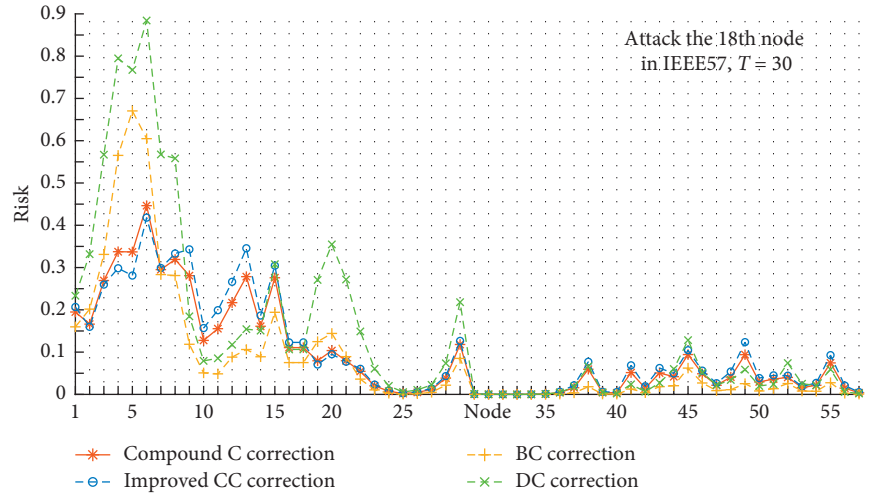


Figure 8: Risk of node under compound C, improved CC, BC, and DC correction at $T = 30$.

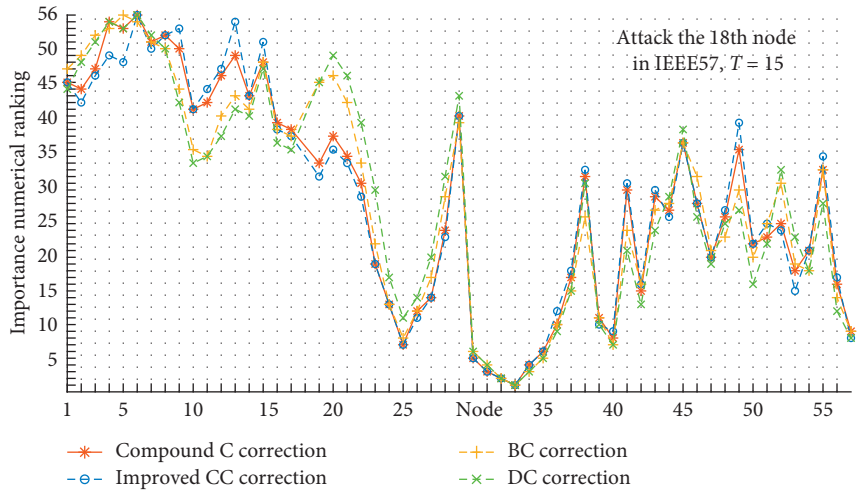


Figure 9: Importance numerical ranking under compound C, improved CC, BC, and DC correction at $T=15$.

Table 3: Compound centrality and S value of node under the IEEE57 attacking the 18th node.

Node	Compound C	S	Node	Compound C	S
1	0.481938925	0.912280702	30	0.310493	0.894736842
2	0.485281429	0.964912281	31	0.260317	0.877192982
3	0.727546619	0.947368421	32	0.267126	0.877192982
4	0.932043014	0.877192982	33	0	0.98245614
5	0.480491466	0.964912281	34	0.303463	0.877192982
6	0.727097074	0.947368421	35	0.351337	0.877192982
7	0.629168326	0.964912281	36	0.408994	0.877192982
8	0.64029582	0.964912281	37	0.460605	0.929824561
9	0.594675807	0.894736842	38	0.546372	0.929824561
10	0.409535734	0.929824561	39	0.308133	0.947368421
11	0.504414267	0.947368421	40	0.309953	0.964912281
12	0.466716998	0.929824561	41	0.436589	0.929824561
13	0.602369165	0.964912281	42	0.127394	0.964912281
14	0.489843246	0.929824561	43	0.170842	0.964912281
15	0.644330887	0.929824561	44	0.435195	0.947368421
16	0.321392772	0.964912281	45	0.494724	0.947368421
17	0.321392772	0.964912281	46	0.376085	0.947368421
18	0.213932971	0.912280702	47	0.331809	0.947368421
19	0.878290586	0.912280702	48	0.381024	0.929824561
20	0.690421858	0.912280702	49	0.526516	0.964912281
21	0.59712313	0.912280702	50	0.349656	0.964912281
22	0.582228702	0.877192982	51	0.307274	0.947368421
23	0.493053805	0.964912281	52	0.399481	0.912280702
24	0.450285109	0.789473684	53	0.342019	0.912280702
25	0.370397806	0.877192982	54	0.377108	0.912280702
26	0.365271476	0.929824561	55	0.458715	0.912280702
27	0.405152159	0.929824561	56	0.379537	0.894736842
28	0.46248975	0.929824561	57	0.24699	0.964912281
29	0.549598151	0.842105263			

the cyber-attack scenario, the minimum risk after the use of the improved CC indicator to modify γ is slightly better than the effect of using the composite centrality indicator to modify.

Figure 8 shows the following:

(1) Node 4 is located in the direct connection position of the infected node 18, and node 6 is in a special position in the connection relationship (with a high BC and DC index). When the conventional centrality is used to modify γ, because the index does not consider the characteristics of each attack scenario, the risk of node 4 and node 6 rises faster, which cannot effectively reduce the risk in the system. Note that the correction effect of the BC is better than that of the DC; in other words, during the attack propagation process, the propagation attribute of the node is more important than the local attribute.

(2) When using the improved CC and compound centrality to modify, since the characteristics of the attack scenario is considered in the index, the attack spread in the system is effectively curbed and the risk of node 4 and node 6 is reduced. Note that the improved CC only considers the actual situation of the attack, which is really effective in reducing the risk of node 4 and node 19 which are close to infected node 18, but does not consider the role of nodes in the subsequent attack propagation. Therefore, the effect of risk reduction for other nodes is worse than that of the compound centrality index.

Figures 7 and 8 show the following:

(1) As the attack spreads, the compound centrality is used to modify γ, which effectively reduces the impact of the attack on the 19th node with a higher initial risk.

(2) Although in the initial stage of attack propagation, the effect of using the improved CC to modify γ is slightly better than that of the compound centrality modification; considering the overall attack propagation, the effect of compound centrality correction γ is better.

In summary, the compound centrality considers both the characteristics of attack scenarios and the role of nodes in

attack propagation, and using it to modify γ can effectively reduce the risk of the remaining nodes in the system. The ranking result of compound centrality is more suitable for identification requirements of key nodes in the power system under cyber-attack scenarios. In the case of $X = 0.03$ and $T = 15\widehat{t}$, the importance ranking of the remaining nodes in the system is shown in Figure 9 when the compound centrality, improved CC, BC, and DC indexes are used to modify γ.

5.2. Cascading Failure Analysis. The node importance evaluation in this paper focuses on the distinction of the importance of a single node. Therefore, according to the cascading failure model established in Section 3, each time a single node is selected for attack and the maximal connected subgraph and the node survival rate S under the stable state are obtained. Analyze the difference of decoupling results when attacking different single nodes. The experiment uses the IEEE57 system "one-to-one correspondence" coupling method. Sequentially attacking a single node of the information network, the power CPS decoupling steady state results are shown in Table 3.

It can be seen that the decoupling caused by the attack on the 24th node is the greatest. After the decoupling, there are only 45 nodes left in the stable system, accounting for 78.9% of the total number of nodes. In the case of cascading failures caused by a cyber-attack on the power CPS, the importance of nodes based on the node survival rate S is ranked as follows: {24,29,4 22 25 31 32 34 35 36 ,9 30 56 , 1 18 19 20 21 52 53 54 55 10 12 13 14 26 27 28 37 38 41 48 , 3 6 11 39 44 45 46 47 51 , 2 5 7 8 13 16 17 23 40 42 43 49 50 57 , 33}. There are some nodes with the same S in the sorting result, so the importance of these nodes cannot be accurately distinguished. The centrality of each node calculated by the compound centrality algorithm of the one-sided network can just make up for this shortcoming. Considering two indexes at the same time, the importance of nodes can be distinguished smoothly.

In summary, the experiment gives indexes that can describe the importance of IEEE57 nodes from two aspects: the compound centrality of the single-layer network and the value S that reflects the degree of cascading failure decoupling, as shown in Table 3. Both indexes can be used as a reference basis for operators and operation centers in the operation of the power system. The dispatch center can select a certain index according to the actual situation of the cyber-attack to distinguish the importance of the node and guide the follow-up strategy.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Key R&D Program of China (2018YFA0702200) and the Fundamental Research Funds for the Central Universities (N2004013).

References

[1] Q. Guo, S. Xin, H. Sun et al., "Power system cyber-physical modelling and security assessment:motivation and ideas," *Proceedings of the CSEE*, vol. 36, no. 6, pp. 1481–1489, 2016.

[2] Y. Liu and P. Ning, *False Data Injection Attacks against State Estimation in Electric Power Grids*, Association for Computing Machinery, New York, NY, USA, 2009.

[3] S. Cui, Z. Han, S. Kar et al., "Coordinated data-injection attack and detection in the smart grid: a detailed look at enriching detection solutions," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 106–115, 2012.

[4] Y. Wang, K. Gao, T. Zhao et al., "Assessing the harmfulness of cascading failures across space in electric cyber-physical system based on improved attack graph," *Proceedings of the CSEE*, vol. 36, no. 6, pp. 1490–1499, 2016.

[5] X. Yi, B. Wang, D. Chen et al., "Review on interdependent networks theory and its applications in the structural vulnerability analysis of electrical cyber-physical system," *Proceedings of the CSEE*, vol. 36, no. 17, pp. 4521–4532, 2016.

[6] A. Vespignani, "The fragility of interdependency," *Nature*, vol. 464, no. 7291, pp. 984-985, 2010.

[7] X. Yi, *Research on Structural Vulnerability of Power Information-Physical System Based on Interdependence Network Theory*, Wuhan University, Wuhan, China, 2016.

[8] Q. Guo, S. Xin, J. Wang et al., "Comprehensive security assessment for a cyber physical energy system a lesson from Ukraine's blackout," *Automation of Electric Power Systems*, vol. 40, no. 5, pp. 1–3, 2016.

[9] Z. Meng, Z. Lu, and J. Song, "Comparison analysis OF the small-world topological model of chinese and american power grids," *Automation of Electric Power Systems*, vol. 28, no. 15, pp. 21–24, 2004.

[10] R. Wang, Q. Sun, P. Zhang et al., "Reduced-order transfer function model of the droop-controlled inverter via Jordan continued-fraction expansion," *IEEE Transactions on Energy Conversion*, vol. 35, no. 3, pp. 1585–1595, 2020.

[11] W. Hu, C. Ruan, H. Nian, and D. Sun, "Zero-sequence current suppression strategy with common mode voltage control for open-end winding PMSM drives with common DC bus," *IEEE Transactions on Industrial Electronics*, vol. 99, 2020.

[12] R. Albert, I. Albert, and G. L. Nakarado, "Structural vulnerability of the North American power grid," *Physical Review E*, vol. 69, no. 2, Article ID 25103, 2004.

[13] M. Kitsak, L. K. Gallos, S. Havlin et al., "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.

[14] S. Dolev, Y. Elovici, and R. Puzis, "Routing betweenness centrality," *Journal of the ACM*, vol. 57, no. 4, pp. 1–27, 2010.

[15] A. N. Langville and C. D. Meyer, "Deeper inside page rank," *Internet Mathematics*, vol. 1, no. 3, 2003.

[16] Y. Chen, A. Hu, and X. Hu, "Evaluation method for node importance in communication networks," *Journal of China Institute of Communications*, vol. 25, no. 8, pp. 129–134, 2004.

[17] Y. Tan, J. Wu, and H. Deng, "Evaluation method for node importance based on node contraction in complex networks," *Systems Engineering Theory& Practice*, vol. 26, no. 11, pp. 79–83, 2006.

[18] Q. Xie, C. Deng, H. Zhao et al., "Evaluation method for node importance of power grid based on the weighted network model," *Automation of Electric Power Systems*, vol. 33, no. 4, pp. 21–24, 2009.

[19] J. Wang, X. Gu, T. Wang et al., "Power system critical node identification based on power tracing and link analysis method," *Power System Protection and Control*, vol. 45, no. 6, pp. 22–29, 2017.

[20] Z. P. Network, "Security situation analysis based on a dynamic bayesian network and phase space reconstruction," *Journal of Supercomputing*, vol. 76, no. 2, pp. 1342–1357, 2020.

[21] G. Li, P. Huang, Y. Chen et al., "Security situation assessment method for cyber physical power system considering threat propagation characteristics," *Electric Power Construction*, vol. 40, no. 5, pp. 29–37, 2019.

[22] V. Buldyrev Sergey, R. Parshani, G. Paul, H. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, 2010.

[23] Y. Tan, X. Li, Y. Cai et al., "Critical node identification for complex power grid based on electrical distance," *Proceedings of the CSEE*, vol. 34, no. 1, pp. 146–152, 2014.

[24] R. Wu, B. Zhang, and L. Tang, "A cascading failure based nodal importance evaluation method applied in dual network coupling model," *Power System Technology*, vol. 4, pp. 1053–1058, 2015.

[25] M. E. J. Newman, "The spread of epidemic disease on networks," *Physical Review E Statal Nonlinear & Soft Matter Physics*, vol. 66, no. 1, Article ID 16128, 2002.

[26] S. Mousavian, J. Valenzuela, and J. Wang, "A probabilistic risk mitigation model for cyber-attacks to PMU networks," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 156–165, 2015.

[27] D. Krackhardt, "Assessing the political landscape: structure, cognition, and power in organizations," *Administrative Science Quarterly*, vol. 35, no. 2, pp. 342–369, 1990.

[28] L. C. Freeman, "Centrality in social networks: conceptual clarification," *Social Networks*, vol. 1, no. 3, 1979.

[29] X. Ji, B. Wang, D. Liu et al., "Improving interdependent networks robustness by adding connectivity links," *Physical A: Statistical Mechanics and its Applications*, vol. 444, pp. 9–19, 2016.

Hindawi
Complexity
Volume 2020, Article ID 6662638, 11 pages
https://doi.org/10.1155/2020/6662638

Research Article

High-Order Observer-Based Sliding Mode Control for the Isolated Microgrid with Cyber Attacks and Physical Uncertainties

Hao Wang,[1] He Jiang,[1] Yan Zhao ,[1] Huanxin Guan,[1] Bo Hu,[2] and Shunjiang Wang[2]

[1]*School of Renewable Energy, Shenyang Institute of Engineering, Shenyang 110136, China*
[2]*State Grid Liaoning Electric Power Supply Co., Ltd., Shenyang 110004, China*

Correspondence should be addressed to Yan Zhao; zhaoyan@sie.edu.cn

Received 22 October 2020; Revised 16 November 2020; Accepted 2 December 2020; Published 15 December 2020

Academic Editor: Rui Wang

System security is essential for the operation of the island microgrid. However, the system security is generally threatened due to the presence of physical uncertainties and cyber attacks. In this article, a novel sliding mode load control strategy is proposed for the microgrid to mitigate cyber attacks and physical uncertainties. Firstly, a high-order disturbance observer (HODO) is designed to estimate the unmeasurable factors in the microgrid. Secondly, a HODO-based sliding mode control (SMC) strategy is proposed where the estimated value observed by the HODO is applied to the sliding mode surface and control law. It can better guarantee the security of the isolated microgrid. Then, the stability of the HODO-based SMC is demonstrated by Lyapunov stability theory. Finally, simulation results show that the proposed control strategy has excellent control performance.

1. Introduction

The power system is regarded as a critical factor for economic development. With the rapid development of communication equipment, power system application, and energy management system [1–3], power systems merging primary and secondary systems are promoted to transform into the cyber-physical system (CPS). Consequently, the security of the cyber-physical power system has received widespread attention.

CPS security includes the security of physical systems and cyber systems. Physical security is the security of the primary system, which can stabilize the system at scheduled operating point under physical uncertainties. Cyber security means the security of the secondary system which is vulnerable to cyber attacks [4–6]. The insecurity of the cyber-physical power system has a significant impact on the society. For example, in 2019, the primary system of Venezuelan power system became the target of an attack, resulting in a large-scale blackout. In 2003, the secondary system, computer network at the power plant, was hacked in Davis–Besse, USA. On the contrary, the security of the power grid can improve the utilization rate of clean energy power generation and enhance the reliability of the power grid [7, 8].

Because of the integration of advanced measuring devices, application software, and renewable generations, the security of the power system is threatened by serious attacks [9–12]. Multitype intelligent analysis software relies on computers and communication networks, which make the system vulnerable to cyber attacks. Meanwhile, the parameters of physical equipment including generators, turbines, and transmission lines are uncertain. Currently, many critical techniques about cyber attacks, which are the major challenge in the CPS, were studied by scholars. There are some advanced resilient control technologies for cyber attacks [13–15], such as data intrusion attacks [16, 17], nontechnical loss fraud, time-delay attacks [18], and replay attacks. Liu and Li [19] proposed a load distribution attack model with the incomplete acquisition of power system information. In [20], a detection technique was studied for uncertain systems. When the power system was attacked, it can be detected and protected immediately. In [21], a control strategy to protect distributed time-delay power systems was proposed. The method of time-delay estimation was introduced to solve time-delay switch attacks. In recent years,

the advanced SMC has been proposed to address the threat of cyber attacks and physical uncertainties [22–24]. In [25], SMC with the neural network observer was constructed, where the measured values were used for the control law, and it was proven to be superior in simulation. Mi et al. [26] proposed SMC based on the proportional-integral sliding mode surface, and this method proved that the microgrid can be immune to the attacks.

In this paper, a control strategy is investigated. Firstly, a transformed dynamics system is established combining cyber attacks and physical uncertainties as a lumped attack. Secondly, the attack is measured by a high-order nonlinear observer where the attack and its derivatives are observed. Then, compared to the linear sliding surface, an improved sliding surface including the estimation value is proposed. By employing the estimation value, system states are forced to move to the sliding mode surface with the control law. Finally, simulation results on the isolated microgrid are carried out to verify the performance of the controller.

The main contributions of the article are as follows:

(1) Considering the characteristic of the power system, the presented HODO can be used to measure the cyber attacks and physical uncertainties in the power system

(2) We construct the sliding mode surface and the control law based on the output of the HODO in the corresponding state space of the microgrid

(3) Using the proposed control strategy, the security of the microgrid will be significantly improved, especially the frequency index

This paper is organized as follows: in Section 2, the system structure of the microgrid and the dynamic equations are proposed. In Section 3, conventional SMC is illustrated. In Section 4, the control strategy is proposed. Firstly, HODO-based SMC is designed. Secondly, the stability is theoretically proved for the proposed method. The experimental simulation results are demonstrated in Section 5, while the work of this paper is summarized in Section 6.

2. Model of the Cyber-Physical Power System

In this paper, a typical cyber-physical system composed of the power system and controller is considered. The matrix form of the cyber-physical power system is expressed as follows:

$$\begin{aligned} \dot{x}(t) &= A_n x(t) + B_n u(t) + F_n \Delta P_d, \\ y &= \overline{C} x(t), \end{aligned} \tag{1}$$

where

$$x(t) = \left[\Delta f(t), \Delta P_g(t), \Delta X_g(t)\right]^T,$$

$$A_n = \begin{bmatrix} -\dfrac{1}{T_P} & \dfrac{K_P}{T_P} & 0 \\ 0 & -\dfrac{1}{T_T} & \dfrac{1}{T_T} \\ -\dfrac{1}{RT_g} & 0 & -\dfrac{1}{T_g} \end{bmatrix},$$

$$B_n = \begin{bmatrix} 0 & 0 & \dfrac{1}{T_g} \end{bmatrix}^T,$$

$$F_n = \begin{bmatrix} -\dfrac{K_P}{T_P} & 0 & 0 \end{bmatrix}^T,$$

$$\overline{C} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T, \tag{2}$$

where $x(t)$ is the system state vector; A_n, B_n, and F_n are system matrices; $\Delta f(t), \Delta P_g(t)$, and $\Delta X_g(t)$ are the deviations of frequency, power output, and governor valve position, respectively; T_p, T_T, and T_g are the time constants of the power system, turbine, and governor, respectively; K_p denotes the power system, R is speed drop; and $u(t)$ and $\Delta P_d(t)$ denote the control vector and the cyber attacks, respectively. The formulated cyber-physical power system is similar to that of the literature [25].

Considering the physical uncertainties of the system dynamic model, equation (1) is written as

$$\begin{aligned} \dot{x}(t) &= \left(A' + \Delta A\right)x(t) + \left(B' + \Delta B\right)u(t) + \left(F' + \Delta F\right)\Delta P_d, \\ y &= \overline{C} x(t), \end{aligned} \tag{3}$$

where A', B', and F' are the determined physical system and ΔA, ΔB, and ΔF denote the uncertainties of the physical system. Equation (4) is the detailed representation of system dynamics (3):

$$\begin{aligned} \Delta \dot{f}(t) &= \left(-\frac{1}{T_p} + k_{11}\right)\Delta f(t) + \left(\frac{K_p}{T_p} + k_{12}\right)\Delta P_g(t) - \left(\frac{K_p}{T_p} + k_{13}\right)\Delta P_d(t), \\ \Delta \dot{P}_g(t) &= \left(-\frac{1}{T_T} + k_{21}\right)\Delta P_g(t) + \left(\frac{1}{T_T} + k_{22}\right)\Delta X_g(t), \\ \Delta \dot{X}_g(t) &= \left(-\frac{1}{RT_g} + k_{31}\right)\Delta f(t) + \left(-\frac{1}{T_g} + k_{32}\right)\Delta X_g(t) + \left(\frac{1}{T_g} + k_{33}\right)u(t), \end{aligned} \tag{4}$$

where k_{mn} denotes physical uncertainties.

The system dynamic model with cyber attacks and physical uncertainties can be represented as follows:

$$\begin{aligned}
\Delta\dot{f}(t) &= -\frac{1}{T_p}\Delta f(t) + \frac{K_p}{T_p}\Delta P_g(t) + d_1(x,t),\\
\Delta\dot{P}_g(t) &= -\frac{1}{T_T}\Delta P_g(t) + \frac{1}{T_T}\Delta X_g(t) + d_2(x,t),\\
\Delta\dot{X}_g(t) &= -\frac{1}{RT_g}\Delta f(t) - \frac{1}{T_g}\Delta X_g(t) + \frac{1}{T_g}u(t) + d_3(x,u,t),
\end{aligned} \tag{5}$$

where

$$\begin{aligned}
d_1(x,t) &= k_{11}\Delta f(t) + k_{12}\Delta P_g(t) - \left(\frac{K_p}{T_p} + k_{13}\right)\Delta P_d(t),\\
d_2(x,t) &= k_{21}\Delta P_g(t) + k_{22}\Delta X_g(t),\\
d_3(x,u,t) &= k_{31}\Delta f(t) + k_{32}\Delta X_g(t) + k_{33}u(t).
\end{aligned} \tag{6}$$

The matrix form of system dynamic model (5) is

$$\begin{aligned}
\dot{x}(t) &= Ax(t) + Bu(t) + F_d d,\\
y &= \overline{C}x(t),
\end{aligned} \tag{7}$$

where $A = A_n, B = B_n, F_d = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, and $d = [d_1, d_2, d_3]^T$.

Assumption 1. Pair A is observable.

In order to design the observer conveniently, let us transform system dynamics (5) using the transformation matrix. The structure is

$$\begin{aligned}
\dot{\eta}_1 &= \eta_2 + d_{11},\\
\dot{\eta}_2 &= \eta_3 + d_{12},\\
\dot{\eta}_3 &= \overline{C}A^3T^{-1}\eta + \overline{C}A^2Bu + d_{13},
\end{aligned} \tag{8}$$

where $\eta = Tx(t)$ and $T = [\overline{C}, \overline{C}A, \overline{C}A^2]^T$.

The aforementioned system is represented as

$$\dot{\eta} = \overline{A}\eta + \overline{B}u(t) + \overline{F}_d d, \tag{9}$$

where $\overline{A} = TAT^{-1}$, $\overline{B} = TB$, $\overline{F}_d = TF_d$, and $[d_{11}, d_{12}, d_{13}]^T = TF_d d$.

For the transformed dynamic equation (9), the following assumption is necessary.

Assumption 2. The attacks d_{1i} are continuous, and their higher-order derivative with respect to time satisfies

$$\left|\frac{\mathrm{d}^q_{1i}(x,t)}{\mathrm{d}t^q}\right| \le \chi, \quad q = 0, 1, 2, 3, \ldots, r,\ i = 1, 2, 3, \tag{10}$$

where χ is a positive number.

Remark 2. Using a linear nonsingular transformation, system dynamic model (5) can be transformed into system (9). It should be noted that system (9) facilitates the design of HODO-based SMC. Meanwhile, η is equivalent to $x(t)$ in simulation analysis.

3. Conventional SMC Design

In the microgrid, the conventional SMC was proposed to ensure system security through secondary frequency regulation of the generator, which adjusts the system to the normal working range with the attacks.

The design of the SMC is composed of two processes: firstly, to design a sliding surface; secondly, to design the control law. The designed sliding surface drives system states to the desired equilibrium asymptotically and remain on it. The system state can be driven to the sliding surface by the designed control law after sufficient time.

3.1. Linear SMC. Based on system dynamic (8), the linear SMC is designed as

$$s_1 = C_1\eta = \sum_{i=1}^{3} c_i\eta_i, \tag{11}$$

where $C_1 = [c_1, c_2, c_3]$ are constants, and $c_3 = 1$. c_i meets that the polynomial $c_1 + c_2p + c_3p^2$, which is Hurwitz, such that the eigenvalues of the polynomial are less than zero.

According to the literature [27], the reaching condition is chosen as $s_1 \cdot \dot{s}_1 < 0$. The equality reaching condition is selected as follows:

$$\dot{s}_1 = -k_{d1}s_1 - k_{\varepsilon1}\mathrm{sign}(s_1), \tag{12}$$

where k_{d1} and $k_{\varepsilon1}$ are positive numbers and $\mathrm{sign}(\cdot)$ is the sign function.

The control law is designed based on (8), (11), and (12), which drives the system state to the sliding surface:

$$u(t) = -(C_1\overline{B})\left[C_1\overline{A}\eta + C_1\chi + k_{d1}s + k_{\varepsilon1}\mathrm{sign}(s)\right]. \tag{13}$$

3.2. Proportional-Integral SMC. The proportional-integral SMC is presented in this section. The proportional-integral sliding surface is selected as

$$s_2 = C_1\eta - \int_0^t (C_1\overline{A} - C_1\overline{B}K)\eta\mathrm{d}t, \tag{14}$$

where matrix K is designed as $\lambda(C_1\overline{A} - C_1\overline{B}K) < 0$.

Similar to (12), we have

$$\dot{s}_2 = -k_{d2}s_2 - k_{\varepsilon2}\mathrm{sign}(s_2). \tag{15}$$

The control law is designed as follows:

$$u(t) = -(C_1\overline{B})\left[C_1\overline{B}K\eta + C_1\chi + k_{d2}s_2 + k_{\varepsilon2}\mathrm{sign}(s_2)\right]. \tag{16}$$

However, there are two obvious drawbacks including large overshoot and the lack of estimation for the attack.

4. Methodology

4.1. High-Order Observer for Cyber Attacks and Physical Uncertainties. In this section, an observer is proposed to estimate the attack [28]. In Figure 1, physical attacks k_{mn} appear in the governing system, turbine, and power system. Meanwhile, cyber attacks $\Delta P_d(t)$ corrupt the power system. When the system is attacked, the boundaries of the undetectable attack will be directly used in SMC without the HODO. Thus, the control is conservative. The proposed control strategy where the HODO can accomplish the detection of unknown attacks compensates this shortcoming to make the controller output more accurate.

The HODO can estimate the attacks for system (9) as follows:

$$\begin{bmatrix} \hat{d}_{11}^{q-1} \\ \hat{d}_{12}^{q-1} \\ \hat{d}_{13}^{q-1} \end{bmatrix} = \begin{bmatrix} p_{11q} \\ p_{12q} \\ p_{13q} \end{bmatrix} + L_q \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}, \tag{17}$$

$$\begin{bmatrix} \dot{p}_{11q} \\ \dot{p}_{12q} \\ \dot{p}_{13q} \end{bmatrix} = -L_q \left[\overline{A}\eta + \overline{B}u(t) + \overline{F}_d \hat{d} \right] + \begin{bmatrix} \hat{d}_{11}^{(q)} \\ \hat{d}_{12}^{(q)} \\ \hat{d}_{13}^{(q)} \end{bmatrix}, \tag{18}$$

$$q = 1, 2, 3, \ldots, r-1,$$

$$\begin{bmatrix} \hat{d}_{11}^{(r-1)} \\ \hat{d}_{12}^{(r-1)} \\ \hat{d}_{13}^{(r-1)} \end{bmatrix} = \begin{bmatrix} p_{11r} \\ p_{12r} \\ p_{13r} \end{bmatrix} + L_q \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}, \tag{19}$$

$$\begin{bmatrix} \dot{p}_{11r} \\ \dot{p}_{12r} \\ \dot{p}_{13r} \end{bmatrix} = -L_r \left[\overline{A}\eta + \overline{B}u(t) + \overline{F}_d \hat{d} \right], \tag{20}$$

where $L_q = \begin{bmatrix} l_{11q} & 0 & 0 \\ 0 & l_{12q} & 0 \\ 0 & 0 & l_{13q} \end{bmatrix}$, $q = 1, 2, 3, \ldots, r$, are constant matrices which are necessary to select $L_q > 0$ for the stability of the HODO; $\tilde{d}_{1i}^{(q-1)}$ and $\hat{d}_{1r}^{(r-1)}$ are estimations of $d_{1i}^{(q-1)}$ and $d_{1r}^{(r-1)}$, respectively, and p_{1iq} are auxiliary variables $(i = 1, 2, 3; q = 1, 2, 3, \ldots, n)$.

The estimation errors are defined as

$$\tilde{e} = \left[\tilde{d}_{11}, \tilde{d}_{12}, \tilde{d}_{13}, \dot{\tilde{d}}_{11}, \dot{\tilde{d}}_{12}, \dot{\tilde{d}}_{13}, \ldots, \tilde{d}_{11}^{r-1}, \tilde{d}_{12}^{r-1}, \tilde{d}_{13}^{r-1} \right], \tag{21}$$

where

$$\begin{aligned} \tilde{d}_{1i} &= d_{1i} - \hat{d}_{1i}, \\ \dot{\tilde{d}}_{1i} &= \dot{d}_{1i} - \dot{\hat{d}}_{1i}, \quad i = 1, 2, 3, \\ \tilde{d}_{1i}^{q-1} &= \frac{d^{q-1} d_{1i}}{dt^{(q-1)}} - \hat{d}_{1i}^{(q-1)}, \quad i = 1, 2, 3; q = 3, 4, 5, \ldots, n, \end{aligned} \tag{22}$$

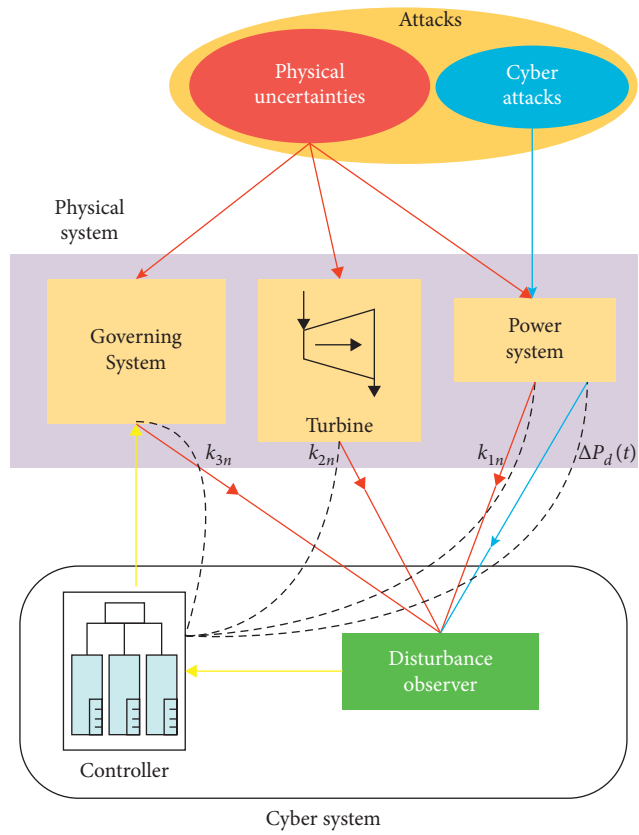


Figure 1: Control system with the observer.

and $\tilde{d}_{1i}^{(q-1)}$ is the error in the estimation of $d_{1i}^{(q-1)}$. From (8), (17), and (18), it follows that

$$\begin{bmatrix} \dot{\tilde{d}}_{11}^{(q-1)} \\ \dot{\tilde{d}}_{12}^{(q-1)} \\ \dot{\tilde{d}}_{13}^{(q-1)} \end{bmatrix} = L_q \begin{bmatrix} \tilde{d}_{11} \\ \tilde{d}_{12} \\ \tilde{d}_{13} \end{bmatrix} + \begin{bmatrix} \hat{d}_{11}^{(q)} \\ \hat{d}_{12}^{(q)} \\ \hat{d}_{13}^{(q)} \end{bmatrix}, \quad q = 1, 2, 3, \ldots, r-1. \tag{23}$$

Obviously, $[\dot{d}_{11}^{q-1}, \dot{d}_{12}^{q-1}, \dot{d}_{13}^{q-1}]^T = [\dot{d}_{11}^{q}, \dot{d}_{12}^{q}, \dot{d}_{13}^{q}]^T$. Subtracting both sides of equation (23) from $[\dot{d}_{11}^{q}, \dot{d}_{12}^{q}, \dot{d}_{13}^{q}]^T$ yields

$$\begin{bmatrix} \dot{d}_{11}^{(q-1)} - \dot{\hat{d}}_{11}^{(q-1)} \\ \dot{d}_{12}^{(q-1)} - \dot{\hat{d}}_{12}^{(q-1)} \\ \dot{d}_{13}^{(q-1)} - \dot{\hat{d}}_{13}^{(q-1)} \end{bmatrix} = -L_q \begin{bmatrix} \tilde{d}_{11} \\ \tilde{d}_{12} \\ \tilde{d}_{13} \end{bmatrix} + \begin{bmatrix} d_{11}^{(q)} - \hat{d}_{11}^{(q)} \\ d_{12}^{(q)} - \hat{d}_{12}^{(q)} \\ d_{13}^{(q)} - \hat{d}_{13}^{(q)} \end{bmatrix}, \tag{24}$$

$$\begin{bmatrix} \dot{\tilde{d}}_{11}^{(q-1)} \\ \dot{\tilde{d}}_{12}^{(q-1)} \\ \dot{\tilde{d}}_{13}^{(q-1)} \end{bmatrix} = -L_q \begin{bmatrix} \tilde{d}_{11} \\ \tilde{d}_{12} \\ \tilde{d}_{13} \end{bmatrix} + \begin{bmatrix} \tilde{d}_{11}^{(q)} \\ \tilde{d}_{12}^{(q)} \\ \tilde{d}_{13}^{(q)} \end{bmatrix}. \tag{25}$$

From (19) and (20), we can get

$$\begin{bmatrix} \dot{\tilde{d}}_{11}^{(r-1)} \\ \dot{\tilde{d}}_{12}^{(r-1)} \\ \dot{\tilde{d}}_{13}^{(r-1)} \end{bmatrix} = -L_r \begin{bmatrix} \tilde{d}_{11} \\ \tilde{d}_{12} \\ \tilde{d}_{13} \end{bmatrix} + \begin{bmatrix} d_{11}^{(r)} \\ d_{12}^{(r)} \\ d_{13}^{(r)} \end{bmatrix}. \tag{26}$$

Differentiating (25) and using (26) give

$$\begin{bmatrix} \tilde{d}_{11}^{(r)} \\ \tilde{d}_{12}^{(r)} \\ \tilde{d}_{13}^{(r)} \end{bmatrix} = -\sum_{q=1}^{r} L_q \begin{bmatrix} \dfrac{d^{(r-q)}\tilde{d}_{11}}{\mathrm{d}t^{(r-q)}} \\ \dfrac{d^{(r-q)}\tilde{d}_{12}}{\mathrm{d}t^{(r-q)}} \\ \dfrac{d^{(r-q)}\tilde{d}_{13}}{\mathrm{d}t^{(r-q)}} \end{bmatrix} + \begin{bmatrix} \dfrac{d^{(r)}d_{11}}{\mathrm{d}t^{r}} \\ \dfrac{d^{(r)}d_{12}}{\mathrm{d}t^{r}} \\ \dfrac{d^{(r)}d_{13}}{\mathrm{d}t^{r}} \end{bmatrix}. \tag{27}$$

Since $[(d_{11}^q(x,t)/\mathrm{d}t^q), (d_{12}^q(x,t)/\mathrm{d}t^q), (d_{13}^q(x,t)/\mathrm{d}t^q)]^T$ is bounded as Assumption 2, the stability of estimation errors depends on the selection of matrices L_q. The HODO error dynamics can be expressed in the matrix form as

$$\dot{\tilde{e}} = D_p\tilde{e} + E\omega, \tag{28}$$

where

$$D_r = \begin{bmatrix} -L_1 & I_3 & 0 & \cdots & 0 \\ -L_2 & 0 & I_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -L_{P-1} & 0 & 0 & \cdots & I_3 \\ -L_P & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$E = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ I_2 \end{bmatrix}, \tag{29}$$

$$\omega = \begin{bmatrix} \dfrac{d^r d_{11}}{\mathrm{d}t^r} \\ \dfrac{d^r d_{12}}{\mathrm{d}t^r} \\ \dfrac{d^r d_{13}}{\mathrm{d}t^r} \end{bmatrix},$$

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In equation (28), the estimation error vector $\tilde{e}$ is illuminated in (21). The derivatives of all vectors in the estimation error vector can be calculated from (25) and (26). From (28), obviously, we can choose appropriate matrix $L_q(q = 1, 2, 3, \ldots, r)$ such that the eigenvalues of D_r can be placed arbitrarily. Assume that L_q are designed to guarantee the eigenvalues of D_r less than zero. A positive symmetric matrix can be selected as follows:

$$D_r^T P + PD_r = -Q. \tag{30}$$

Define a Lyapunov functional, and $\lambda_{\min}$ is the smallest eigenvalue; then,

$$V(\tilde{e}) = \tilde{e}^T P\tilde{e}. \tag{31}$$

Substituting (28) into the derivative of $V(\tilde{e})$ becomes

$$\begin{aligned} \dot{V}(\tilde{e}) &= \dot{\tilde{e}}^T P\tilde{e} + \tilde{e}^T P\dot{\tilde{e}} \\ &= \dot{\tilde{e}}^T \left(D_p^T P + PD_p\right)\tilde{e} + 2\tilde{e}^T PE\omega \\ &\le -\tilde{e}^T Q\tilde{e} + 2\|PE\|\|\tilde{e}\|\omega\| \\ &\le -\lambda_{\min}\|\tilde{e}\|^2 + 2\|PE\|\|\tilde{e}\|\omega\| \\ &= -\|\tilde{e}\|\left(\lambda_{\min}\|\tilde{e}\| - 2\|PE\|\|\omega\|\right). \end{aligned} \tag{32}$$

Consequently, for (8), (25), and (26), after sufficiently long time, the norm of the estimation error is ultimately bounded by

$$\|\tilde{e}\| \le \frac{2\|PE\|\|\omega\|}{\lambda_{\min}}. \tag{33}$$

When the error state trajectory enters into the closed ball centered at $\tilde{e} = 0$ with radius φ and the smallest eigenvalue $\lambda_{\min} > 0$, the Lyapunov function satisfies $\dot{V}(\tilde{e}) < 0$. It implies that the estimation error system is stable. The bound of the estimation error can be lowered by the appropriate choice of the parameter $L_q(q = 1, 2, 3, \ldots, r)$.

4.2. SMC Based on the HODO. When the system is attacked, SMC is an effective control strategy to guarantee the security of the system. However, the shortcoming is that the conventional SMC method would bring some adverse effects such as introducing overshoot. In this section, a HODO-based SMC is presented. It should be noted that SMC and HODO are designed, respectively.

To be immune to attacks, a linear sliding surface based on the HODO is applied to improve the stability of the power system.

The sliding surface is selected as follows:

$$s^* = \sum_{i=1}^{3} c_i\eta_i + c_2\hat{\tilde{d}}_{11} + \hat{\dot{\tilde{d}}}_{11} + \hat{\tilde{d}}_{12}. \tag{34}$$

Theorem 1. *Using HODO (17)–(20) and the designed controller law (35), all the states of (9) are ultimately bounded; therefore, the closed-loop system is asymptotically stable around equilibrium with the following control law:*

$$u = -\frac{1}{CA^2B}\left[c_1\left(\eta_2 + \hat{d}_{11}\right) + c_2\left(\eta_3 + \hat{d}_{12}\right) + \overline{C}A^3T^{-1}\eta + \hat{d}_{13} + \varsigma\right], \tag{35}$$

where

$$\varsigma = k_d s^* + k_\varepsilon sign(s^*),$$
$$k_\varepsilon = \left\{\left[(c_1 + c_2 l_{111} + l_{112}) + (c_2 + l_{121}) + 4\right]\left\|\tilde{e}_1\right\| + 3\chi + \theta\right\}, \tag{36}$$

and k_d and θ are positive constants.

Proof. Construct a Lyapunov candidate function as

$$V(s^*) = \frac{1}{2}s^{*2}. \tag{37}$$

It obviously elicits

$$\dot{V}(s^*) = s^*\dot{s}^*. \tag{38}$$

From (9) and (34), it follows that

$$\dot{s}^* = c_1\dot{\eta}_1 + c_2\dot{\eta}_2 + \dot{\eta}_3 + c_2\dot{\hat{d}}_{11} + \dot{\hat{\hat{d}}}_{11} + \dot{\hat{d}}_{12} - \zeta. \tag{39}$$

Inserting (35) into (39), we obtain

$$\begin{aligned}\dot{s}^* = {} & (c_1 + c_2 l_{111} + l_{112})\dot{\tilde{d}}_{11} + (c_2 + l_{121})\dot{\tilde{d}}_{12} \\ & + \dot{\tilde{d}}_{13} + \dot{\hat{d}}_{11} + \dot{\hat{d}}_{12} + \dot{\hat{\hat{d}}}_{11} - \zeta.\end{aligned} \tag{40}$$

According to (24), we have

$$\begin{cases}\dot{\hat{d}}_{11} = l_{111}\tilde{d}_{11} + \hat{\dot{d}}_{11} \\ \dot{\hat{d}}_{12} = l_{121}\tilde{d}_{12} + \hat{\dot{d}}_{12} \\ \dot{\hat{d}}_{13} = l_{131}\tilde{d}_{13} + \hat{\dot{d}}_{13} \\ \dot{\hat{\hat{d}}}_{11} = l_{112}\tilde{d}_{11} + \hat{\ddot{d}}_{11} \\ \dot{\hat{\hat{d}}}_{12} = l_{122}\tilde{d}_{12} + \hat{\ddot{d}}_{12} \\ \dot{\hat{\hat{d}}}_{13} = l_{132}\tilde{d}_{13} + \hat{\ddot{d}}_{13}.\end{cases} \tag{41}$$

Then, substituting (41) to (40), we have

$$\begin{aligned}\dot{s}^* = {} & (c_1 + c_2 l_{111} + l_{112})\dot{\tilde{d}}_{11} + (c_2 + l_{121})\dot{\tilde{d}}_{12} + \dot{\tilde{d}}_{13} \\ & + \dot{d}_{11} - \dot{\hat{d}}_{11} + \dot{d}_{12} - \dot{\hat{d}}_{12} + \ddot{d}_{11} - \dot{\hat{\hat{d}}}_{11} - \zeta.\end{aligned} \tag{42}$$

Substituting (42) into (38), it follows that

$$\begin{aligned}\dot{V}(s^*) = s^*\dot{s}^* \le {} & -k_\varepsilon\left|s^*\right| - k_d s^{*2} + [\,(c_1 + c_2 l_{111} + l_{112}) \\ & + (c_2 + l_{121}) + 4\,]\left\|\tilde{e}_1\right\|\left|s^*\right| + 3\chi\left|s^*\right| \\ < {} & -\theta\left|s^*\right| - k_d s^{*2} < 0.\end{aligned} \tag{43}$$

According to (21), $\tilde{e}_1 = [\tilde{d}_{11}, \tilde{d}_{12}, \tilde{d}_{13}, \dot{\tilde{d}}_{11}, \dot{\tilde{d}}_{12}, \dot{\tilde{d}}_{11}]$ is bounded as follows:

$$\left\|\tilde{e}_1\right\| < \|\tilde{e}\| \le \frac{2\|PE\|\|\omega\|}{\lambda_m}. \tag{44}$$

This completes the proof.

It should be noted that CA^2B is invertible. The control block diagram of the proposed HODO-based SMC is shown in Figure 2.

Figure 2 briefly illuminates the steps of the HODO-based SMC. Firstly, the system dynamics is obtained, and coordinate transformation is performed. Secondly, the transformed coordinate is applied in the HODO. Thirdly, the estimated values calculated by the HODO work on the sliding mode surface. Finally, the calculated control law obtained from the sliding mode surface has an effect on the power system and the HODO. □

5. Simulation Results with SMC

In this section, the proposed control strategy is analyzed by several numerical simulations. As we all know, in the cyber-physical power system, frequency deviation is the most important security index. Consequently, the frequency deviation is mainly discussed in the simulation.

In order to verify the effectiveness of the studied control strategy for the CPS, comparative results between the HODO-based SMC and the conventional SMC are simulated and analyzed. The system parameters for simulation are listed in Table 1 [23]. In this simulation, the second- and the third-order observer are analyzed for the effect of this control strategy.

By using the HODO based on equations (17)–(20), the second- and third-order disturbance observers are expressed as follows.

The second-order disturbance observer:

$$\begin{aligned}\hat{d}_{1i} &= p_{1i1} + l_{1i1}\eta_i, \\ \dot{p}_{1i1} &= -l_{1i1}\left(\eta_{i+1} + \hat{d}_{1i}\right) + \hat{\dot{d}}_{1i}, \\ \hat{\dot{d}}_{1i} &= p_{1i2} + l_{1i2}\eta_i, \\ \dot{p}_{1i2} &= l_{1i2}\left(\eta_{i+1} + \hat{d}_{1i}\right), \\ \hat{d}_{13} &= p_{131} + l_{131}\eta_3, \\ \dot{p}_{131} &= -l_{131}\left(-77.4\eta_1 - 42.8\eta_2 - 17\eta_3 + 189.4u + \hat{d}_{13}\right) + \hat{\dot{d}}_{13}, \\ \hat{\dot{d}}_{13} &= p_{132} + l_{132}\eta_3, \\ \dot{p}_{132} &= l_{132}\left(-77.4\eta_1 - 42.8\eta_2 - 17\eta_3 + 189.4u + \hat{d}_{13}\right).\end{aligned} \tag{45}$$

The third-order disturbance observer:

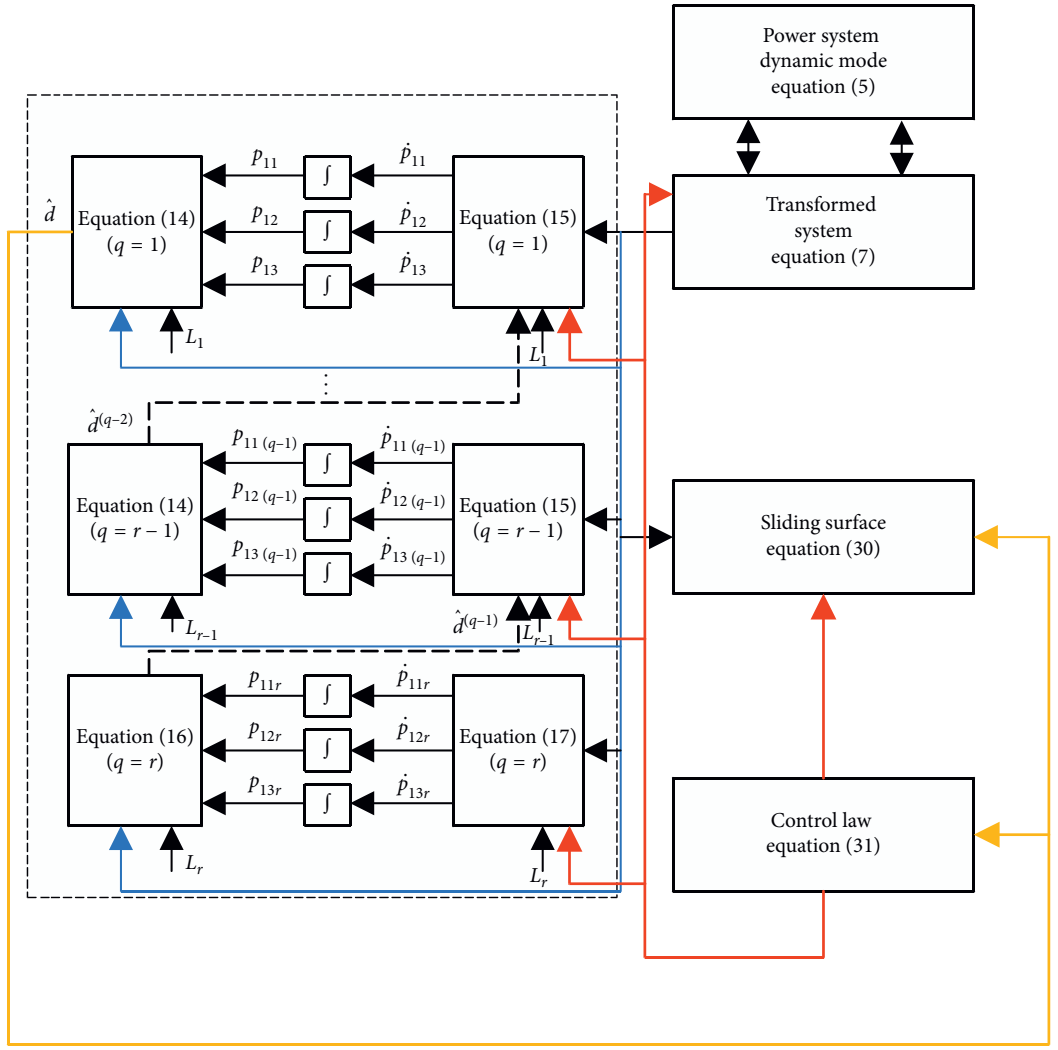


Figure 2: Block diagram of the system with the proposed SMC.

Table 1: The parameters of power systems.

Parameters	R	K_P	T_P	T_T	T_G
Values	2.7	112.5	25	0.33	0.072

$$
\begin{aligned}
\hat{d}_{13} &= p_{131} + l_{131}\eta_3,\\
\dot{p}_{131} &= -l_{131}\left(-77.4\eta_1 - 42.8\eta_2 - 17\eta_3 + 189.4u + \hat{d}_{13}\right) + \dot{\hat{d}}_{13},\\
\dot{\hat{d}}_{13} &= p_{132} + l_{132}\eta_3,\\
\dot{p}_{132} &= l_{132}\left(-77.4\eta_1 - 42.8\eta_2 - 17\eta_3 + 189.4u + \hat{d}_{13}\right) + \ddot{\hat{d}}_{13}\\
\ddot{\hat{d}}_{13} &= p_{133} + l_{133}\eta_3,\\
\dot{p}_{133} &= l_{133} - 77.4\eta_1 - 42.8\eta_2 - 17\eta_3 + 189.4u + \hat{d}_{13}.
\end{aligned}
\tag{46}
$$

The control parameters and the initial variables are selected as

$$
\begin{aligned}
c_1 &= 8,\\
c_2 &= 3,\\
c_3 &= 1,\\
k_d &= k_{d1}\\
k_\varepsilon &= k_{\varepsilon 1}\\
x(0) &= [0.1, 0, 0]^T.
\end{aligned}
\tag{47}
$$

5.1. Step Cyber Attack. In this case, a step cyber attack is applied without physical uncertainties. Cyber attack is executed to the microgrid, which is 0.1 pu. And the cyber attack is added at the initial time and ends at 5 s. The attack boundary is $\chi = 0.1$.

The values are considered as follows:

Parameter 1 (P1): $l_{111} = l_{121} = l_{131} = 10000$ and $l_{112} = l_{122} = l_{132} = 1500$

Parameter 2 (P2): $l_{111} = l_{121} = l_{131} = 10000$ and $l_{112} = l_{122} = l_{132} = 1000$

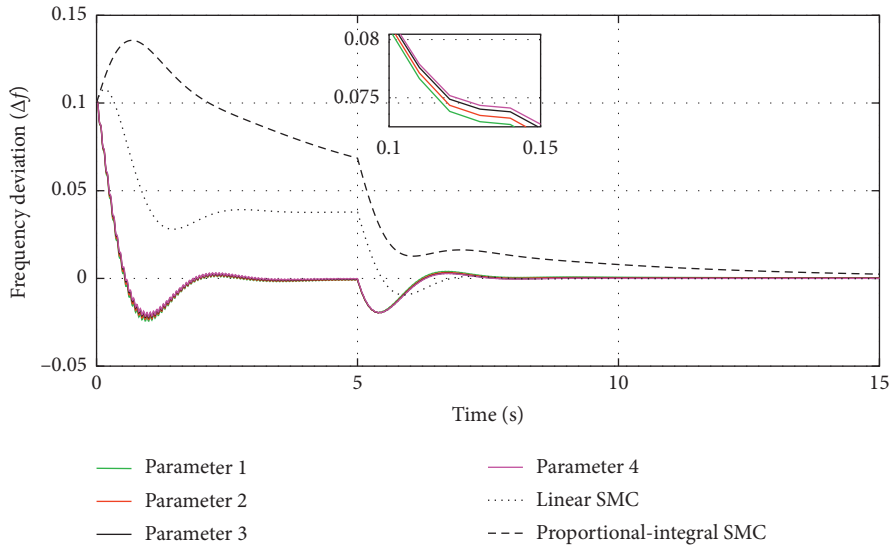


Figure 3: Frequency deviation with the step load attack.

Parameter 3 (P3): $l_{111} = l_{121} = l_{131} = 10000$ and $l_{112} = l_{122} = l_{132} = 500$

Parameter 4 (P4): $l_{111} = l_{121} = l_{131} = 10000$ and $l_{112} = l_{122} = l_{132} = 200$

We can see that the proposed SMC can ensure system security (especially, frequency deviation). The simulation result is presented in Figure 3. Moreover, it has a small overshoot compared with the conventional SMC at the initial time.

An evaluation index based on frequency deviation is employed to demonstrate HODO's control performance, which is

$$f_D = \int_0^t |\Delta f| \mathrm{d}t. \tag{48}$$

Table 2 shows the evaluation index of frequency deviation between the conventional SMC and the second-order disturbance observer with different parameters. It can be concluded from Table 2 that the adjustment ability of the second-order disturbance observer is superior to the conventional SMC. Furthermore, the selection of the parameter in the second-order disturbance observer has an improvement on the overshoot.

5.2. Random Cyber Attacks. In this case, the designed SMC with the second-order disturbance observer and third-order disturbance observer is tested in the microgrid. The random external attack is injected into the microgrid, which is $\Delta P_d(t) = 0.2 \sin(t)$ and ends at 15 s. The parameters of the disturbance observer are as follows:

(a) The second-order disturbance observer:

$$\begin{aligned} l_{111} &= l_{121} = l_{131} = 500, \\ l_{112} &= l_{122} = l_{132} = 100. \end{aligned} \tag{49}$$

(b) The third-order disturbance observer:

Table 2: The evaluation index of frequency deviation.

	Linear SMC	Proportional-integral SMC	P1	P2	P3	P4
f_D	0.2423	0.5977	0.0661	0.0642	0.0625	0.0617

$$\begin{aligned} l_{111} &= l_{121} = l_{131} = 500, \\ l_{112} &= l_{122} = l_{132} = 100, \\ l_{113} &= l_{123} = l_{133} = 25. \end{aligned} \tag{50}$$

The plot of the estimation values $\Delta \widehat{P}_d(t)$ is shown in Figure 4. The disturbance can be tracked in 8 s, while the estimated value accurately estimates the reference disturbance after 10 s. The frequency deviation with the HODO-based SMC is presented in Figure 5.

Using equation (48), we get

$$\begin{aligned} f_{D2} &= \int_{20}^{22} |\Delta f| \mathrm{d}t = 0.007, \\ f_{D3} &= \int_{20}^{22} |\Delta f| \mathrm{d}t = 0.003, \end{aligned} \tag{51}$$

where f_{D2} and f_{D3} represent the evaluation index of the second- and third-order disturbance observer.

It can be concluded that it quickly converges to 0 with the third-order disturbance observer, compared with the effect of the second-order disturbance observer.

5.3. Physical Uncertainties with Cyber Attacks. The stochastic step attack is executed to the microgrid (see Figure 6). Adjust the uncertainty time constant T_T, the governor time constant T_g, and the speed drop R to 0.38, 0.08, and 2.5 in 20 s, respectively. System uncertainties and cyber attacks occur simultaneously between 15 and 20 seconds. When physical uncertainties appear in the secondary system, the system variables of the microgrid are gradually stable within the limited time by using HODO-based SMC, which are shown in Figure 7.

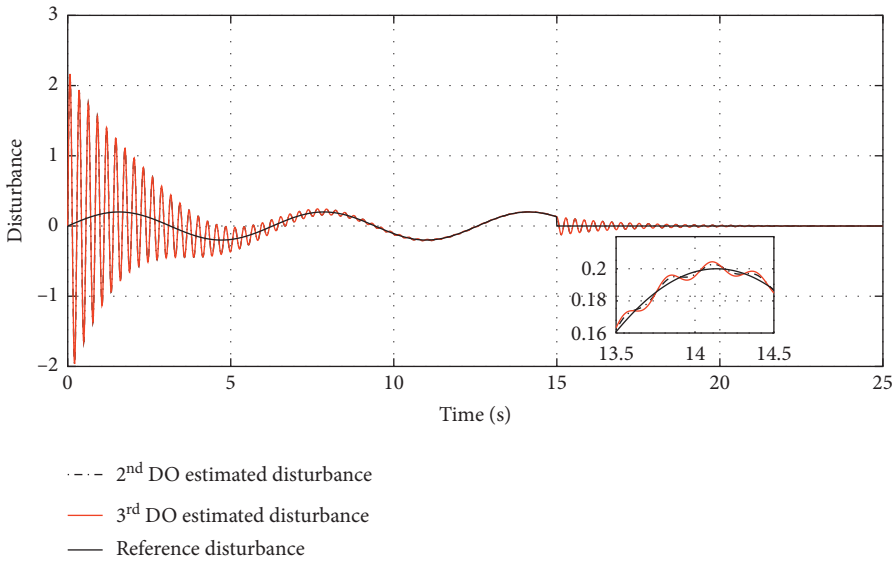


Figure 4: Actual and estimated attack.

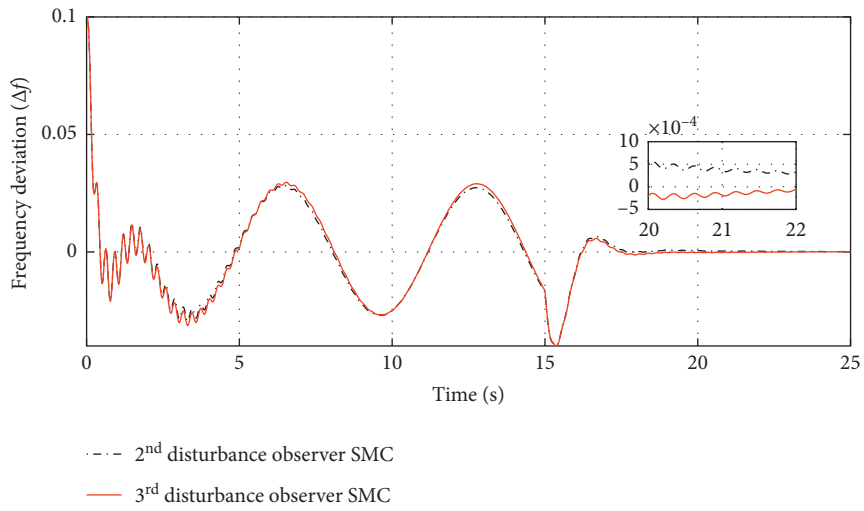


Figure 5: Frequency deviation with the disturbance observer.

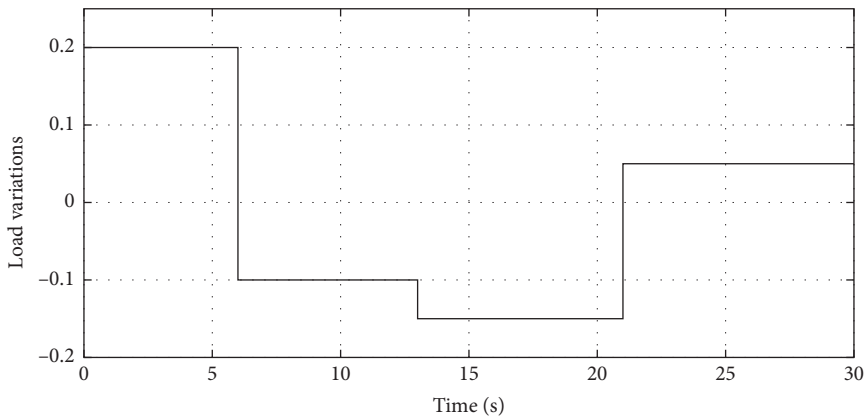


Figure 6: Stochastic attack.

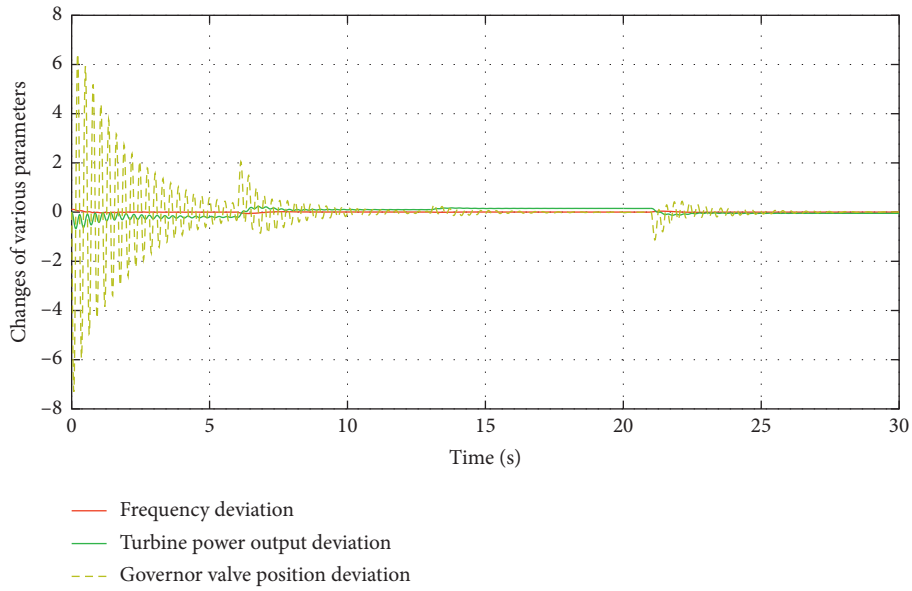


Figure 7: State deviation (frequency, power output, and governor valve position) of the microgrid.

6. Conclusion

In this paper, a HODO-based SMC is employed to guarantee the security of the cyber-physical power system. Firstly, HODO is applied to measure cyber attacks and physical uncertainties with matching and unmatching. Secondly, the SMC with the estimated value obtained by the HODO effectively stabilizes the system, furthermore, as compared with the conventional SMC and the proposed control strategy, and the advantage of the HODO-based SMC is small overshoot. In future, further research will be extended to the power-interconnected and time-delay system.

Data Availability

The data used in the research of this article are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Liaoning Revitalization Talents Program (XLYC1907138), the Natural Science Foundation of Liaoning Province (2019-MS-239), the Doctoral Scientific Research Foundation of Liaoning Province (2020-BS-181), the Key R&D Program of Liaoning Province (2018220017), and the Technology Innovation Talent Fund of Shenyang (RC190360).

References

[1] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber-physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, 2012.

[2] M. Govindarasu, A. Hann, and P. Sauer, "Cyber-physical systems security for smart grid," in *Future Grid Initiative White Paper*PSERC Arizona State University, Tempe, AZ, USA, 2012.

[3] Q. Sun, Y. Zhang, H. He, D. Ma, and H. Zhang, "A novel energy function-based stability evaluation and nonlinear control approach for energy internet," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1195–1210, 2017.

[4] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, 2017.

[5] B. M. Horowitz and K. M. Pierce, "The integration of diversely redundant designs, dynamic system models, and state estimation technology to the cyber security of physical systems," *Systems Engineering*, vol. 16, no. 4, pp. 401–412, 2013.

[6] Z. Qu, Y. Zhang, N. Qu, L. Wang, Y. Li, and Y. Dong, "Method for quantitative estimation of the risk propagation threshold in electric power CPS based on seepage probability," *IEEE Access*, vol. 6, pp. 68813–68823, 2018.

[7] Q. Sun, R. Han, H. Zhang, J. Zhou, and J. M. Guerrero, "A multiagent-based consensus algorithm for distributed coordinated control of distributed generators in the energy internet," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 3006–3019, 2015.

[8] A. Giani, E. Bitar, M. Garcia, M. Mcqueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1244–1253, 2013.

[9] J. P. Lopes, N. Hatziargyriou, J. Mutale, P. Djapic, and N. Jenkins, "Integrating distributed generation into electric power systems: a review of drivers, challenges and opportunities," *Electric Power Systems Research*, vol. 77, no. 5, pp. 1189–1203, 2007.

[10] Y. Li, H. Zhang, X. Liang, and B. Huang, "Event-triggered-based distributed cooperative energy management for multienergy systems," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2008–2022, 2019.

[11] F. Blaabjerg, R. Teodorescu, M. Liserre, and A. V. Timbus, "Overview of control and grid synchronization for distributed

power generation systems," *IEEE Transactions on Industrial Electronics*, vol. 53, no. 5, pp. 1398–1409, 2006.

[12] R. Wang, Q. Sun, D. Ma, and Z. Liu, "The small-signal stability analysis of the droop-controlled converter in electromagnetic timescale," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1459–1469, 2019.

[13] L. An and G.-H. Yang, "Improved adaptive resilient control against sensor and actuator attacks," *Information Sciences*, vol. 423, pp. 145–156, 2018.

[14] W. Ao, Y. Song, and C. Wen, "Adaptive cyber-physical system attack detection and reconstruction with application to power systems," *IET Control Theory & Applications*, vol. 10, no. 12, pp. 1458–1468, 2016.

[15] Y. Wu, Z. Wei, J. Weng, X. Li, and R. H. Deng, "Resonance attacks on load frequency control of smart grids," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4490–4502, 2018.

[16] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 Ukraine blackout: implications for false data injection attacks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3317-3318, 2017.

[17] Y. Chen, S. Huang, F. Liu, Z. Wang, and X. Sun, "Evaluation of reinforcement learning-based false data injection attack to automatic voltage control," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2158–2169, 2019.

[18] A. Sargolzaei, K. K. Yen, and M. N. Abdelghani, "Control of nonlinear heartbeat models under time-delay-switched feedback using emotional learning control," *International Journal on Recent Trends in Engineering & Technology*, vol. 10, no. 2, pp. 85–91, 2014.

[19] X. Liu and Z. Li, "Local load redistribution attacks in power systems with incomplete network information," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1665–1676, 2014.

[20] H. Wang, J. Ruan, B. Zhou et al., "Dynamic data injection attack detection of cyber physical power systems with uncertainties," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5505–5518, 2019.

[21] A. Sargolzaei, K. K. Yen, and M. N. Abdelghani, "Preventing time-delay switch attack on load frequency control in distributed power systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1176–1185, 2016.

[22] L. Jiang, W. Yao, Q. H. Wu, J. Y. Wen, and S. J. Cheng, "Delay-dependent stability for load frequency control with constant and time-varying delays," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 932–941, 2012.

[23] A. Y. Siaramakrishna, M. V. Hariharam, and M. C. Srisailam, "Design of variable structure load-frequency controller using pole assignment technique," *International Journal of Control*, vol. 40, no. 3, pp. 487–498, 1984.

[24] C.-T. Pan and C.-M. Liaw, "An adaptive controller for power system load-frequency control," *IEEE Transactions on Power Systems*, vol. 4, no. 1, pp. 122–128, 1989.

[25] C. Mu, W. Liu, W. Xu, and M. Rabiul Islam, "Observer-based load frequency control for island microgrid with photovoltaic power," *International Journal of Photoenergy*, vol. 2017, Article ID 2851436, 11 pages, 2017.

[26] Y. Mi, Y. Fu, C. Wang, and P. Wang, "Decentralized sliding mode load frequency control for multi-area power systems," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4301–4309, 2013.

[27] V. I. Utkin, *Sliding Modes in Control Optimization*, Springer, Berlin, Germany, 1992.

[28] D. Ginoya, P. D. Shendge, and S. B. Phadke, "Sliding mode control for mismatched uncertain systems using an extended disturbance observer," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 4, pp. 1983–1992, 2014.

Hindawi
Complexity
Volume 2020, Article ID 6641758, 12 pages
https://doi.org/10.1155/2020/6641758

WILEY | Hindawi

Research Article

Disturbance Observer-Based Adaptive Neural Network Control of Marine Vessel Systems with Time-Varying Output Constraints

Wei Zhao,[1] **Li Tang**,[2] **and Yan-Jun Liu**[2]

[1]*School of Electrical Engineering, Liaoning University of Technology, Jinzhou, Liaoning 121001, China*
[2]*College of Science, Liaoning University of Technology, Jinzhou, Liaoning 121001, China*

Correspondence should be addressed to Li Tang; tangli413@hotmail.com

Received 10 October 2020; Revised 30 October 2020; Accepted 7 November 2020; Published 25 November 2020

Academic Editor: Qiuye Sun

This article investigates an adaptive neural network (NN) control algorithm for marine surface vessels with time-varying output constraints and unknown external disturbances. The nonlinear state-dependent transformation (NSDT) is introduced to eliminate the feasibility conditions of virtual controller. Moreover, the barrier Lyapunov function (BLF) is used to achieve time-varying output constraints. As an important approximation tool, the NN is employed to approximate uncertain and continuous functions. Subsequently, the disturbance observer is structured to observe time-varying constraints and unknown external disturbances. The novel strategy can guarantee that all signals in the closed-loop system are semiglobally uniformly ultimately bounded (SGUUB). Finally, the simulation results verify the benefit of the proposed method.

1. Introduction

In recent years, the adaptive control [1, 2] of nonlinear systems [3–7] is attracted more and more attention. As two main approximation tools, neural networks (NNs) and fuzzy logic systems (FLSs) [8, 9] are used to dispose some uncertain variables and unknown functions. For example, in [10], by using NNs, the neural network adaptive fault-tolerant control method under multiobjective constraints is designed. An effective finite-frequency $H-/H\infty$ fault detection method is proposed for the descriptor system in [11]. It is noted that the aforementioned studies neglect the influence of constraints on nonlinear systems.

In general, because of physical limitations and safety factors of the system, multifarious constraints exist in most practical systems, such as flexible string system [12], flexible-joint robot manipulators [13], nonuniform gantry crane systems [14], and flexible aerial refueling hose [15]. The barrier Lyapunov function (BLF) is presented in [16] for strict feedback systems. Since then, as a main tool, the BLF plays a crucial role in dealing with constraints. Subsequently, full state constraints adaptive controller is designed based on BLF and backstepping technique in [17]. In order to handle the control problems of state-constrained nonlinear systems better, in [18, 19], the authors present an integral BLF. It is worth noting that none of the above articles involve time-varying constraints. In [20–22], the adaptive controller of time-varying full-state constraint is constructed. Thus, the constant constraints expand the time-varying constraints. However, there is a limitation in the adaptive control design based on BLF, i.e., the feasibility of the virtual controller, which will increase the cost of control design. Hence, how to eliminate the feasibility of the controller is an urgent problem. It is noted that few of them involve the feasibility conditions of intermediate controllers to marine vessel systems, which is also a challenge of this paper.

More recently, artificial intelligence has made impressive progress. More and more adaptive tracking control of marine vessel is widely studied. Subsequently, a number of significant results are proposed around the adaptive control design of the ship. The adaptive NN controller for ships is designed in [23]. Afterward, much effort has been made in addressing the situation where the ship system parameters are unknown. For the sliding mode method based on the model, an adaptive control approach is proposed in [24]. The precise recognition and learning control of marine ships

under an unknown dynamic environment is further developed in [25]. Based on the above description, we propose an adaptive NN control method of marine vessel with time-varying output constraints.

In this article, we try our best to propose an adaptive NN control method for the marine vessel system with disturbance observer and time-varying output constraints. Furthermore, NNs are developed to deal with uncertain parameters and unknown functions. As is known to all, in the existing literature, this kind of marine vessel system with time-varying state constraints is rarely dealt with by removing the feasibility of virtual controller. A novel transformation function is proposed to ensure that the states do not violate the bounds. The main work and contributions of this article are as follows:

(1) As is known to all, most of the existing constraint control methods involve feasibility conditions. In this article, a new coordinate transformation is designed to resolve the time-varying position constraint of marine surface vessel without involving feasibility conditions. The coordinate transformation is used to get a control scheme that can avoid the feasibility condition, which is conducive to the realization of the control scheme.

(2) An exponential convergence disturbance observer is designed to deal with time-varying constraints and uncertain external disturbances, which are created by wind, waves, currents, and so on.

The specific organization of this paper is as follows. First, the dynamic model of the surface vessels is introduced. Then, the detailed procedure of BLF derivation for dealing with time-varying output constraints is given, in which the disturbance observer, the adaptive controller, and the adaptive laws are designed. Finally, simulation results verify the effectiveness of the proposed method.

2. Preliminaries

We consider the dynamics of 3 degree-of-freedom (3DOF) surface ships with model uncertainty and external disturbance. The dynamic model between the vessel's position and velocity can be expressed as follows:

$$\dot{\eta} = R(\eta)\nu, \tag{1}$$

where the vector $\eta = [\eta_x, \eta_y, \eta_\psi]^T \in R^3$ stands for the output of earth-frame positions and heading. $[\eta_x, \eta_y] \in R^2$ is the position of the vessel, and $\eta_\psi \in R$ denotes the heading, respectively. The vector $\nu = [\nu_x, \nu_y, \nu_\psi]^T \in R^3$ represents the velocity vector of the vessel. The rotation matrix $R(\eta)$ can be described as follows:

$$\begin{aligned} R(\eta) &= \begin{bmatrix} \cos\eta_\psi & -\sin\eta_\psi & 0 \\ \sin\eta_\psi & \cos\eta_\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\ R^{-1}(\eta) &= R^T(\eta). \end{aligned} \tag{2}$$

In this paper, the output of position and heading of marine vessel system is constrained by

$$-F_{k1}(t) \le \eta \le F_{k2}(t), \quad k = 1, 2, 3, \tag{3}$$

where

$$\begin{aligned} F_{k1}(t) &= \left[F_{11}(t), F_{21}(t), F_{31}(t)\right]^T, \\ F_{k2}(t) &= \left[F_{12}(t), F_{22}(t), F_{32}(t)\right]^T. \end{aligned} \tag{4}$$

The relationship between the vessel's velocity and the force acting on the vessel can be described by the dynamic model of vessel [26] as follows:

$$M\dot{\nu} + A(\nu)\nu + B(\nu)\nu + f(\eta) = \tau + g, \tag{5}$$

where $g(t) = [g_1(t), g_2(t), g_3(t)]^T$ is unknown and time-varying external disturbance caused by wind and waves, and $f(\eta)$ is unknown restorative force caused by the ship's gravity and buoyancy. The control input is $\tau = [\tau_1, \tau_2, \tau_3]^T$. Besides, $M = M^T \in R^{3\times3}$ denotes the positive definite inertia matrix, $A(\nu) \in R^{3\times3}$ represents the torques of Coriolis and centripetal, and $B(\nu) \in R^{3\times3}$ stands for damping matrix. They are, respectively,

$$\begin{aligned} M &= \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix}, \\ A(\nu) &= \begin{bmatrix} 0 & 0 & -a_{22}\nu_y - a_{23}\nu_\psi \\ 0 & 0 & a_{11}\nu_x \\ a_{22}\nu_y + a_{23}\nu_\psi & -a_{11}\nu_x & 0 \end{bmatrix}, \\ B(\nu) &= \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & b_{23} \\ 0 & b_{32} & b_{33} \end{bmatrix}, \end{aligned} \tag{6}$$

where

$$\begin{aligned} b_{11} &= -X_1 - X_{11}|\nu_x| - X_{111}\nu_x^2, \\ b_{22} &= -Y_2 - Y_{22}|\nu_y| - Y_{32}|\nu_\psi|, \\ b_{23} &= -Y_3 - Y_{23}|\nu_y| - Y_{33}|\nu_\psi|, \\ b_{32} &= -N_2 - N_{22}|\nu_y| - N_{32}|\nu_\psi|, \\ b_{33} &= -N_3 - N_{23}|\nu_y| - N_{33}|\nu_\psi|, \end{aligned} \tag{7}$$

with $a_{11}, a_{22}, a_{23}, a_{32}, a_{33}, X_1, X_{11}, X_{111}, Y_2, Y_{22}, Y_{32}, Y_3, Y_{23}, Y_{33}, N_2, N_{22}, N_{32}, N_3, N_{23}$, and N_{33} are design parameters.

Define $x_1 = \eta$ and $x_2 = \nu$, and then (1) and (5) can be transformed into the following equation:

$$\begin{cases} \dot{x}_1 = R(x_1)x_2, \\ \dot{x}_2 = M^{-1}\left(\tau + g - f(x_1) - A(x_2)x_2 - B(x_2)x_2\right). \end{cases} \tag{8}$$

The main task of this article is to design controller and adaptive laws so that the system output of vessel $x_1 = [x_{11}, x_{12}, x_{13}]^T$ tracks the expected trajectory $\varpi_d = [\varpi_{d1}, \varpi_{d2}, \varpi_d]^T$ as effectively as possible. Meanwhile, all

signals are bounded in the closed-loop systems and their constraints are not violated forever.

As important approximation methods in nonlinear adaptive control, the radial basis function neural networks (RBFNNs) are usually utilized to deal with unknown and continuous functions. In this article, the RBFNNs are applied to approximate the unknown and continuous function $F(Z)$. The following relation holds:

$$F(Z) = \theta^T \varphi(Z) + \varsigma(Z), \tag{9}$$

where $\varsigma(Z)$ represents approximation error and satisfies $\varsigma(Z) \leq \bar{\varsigma}$, which will be used in the following derivation. $Z \in \Omega_Z \subset R^q$ is the input vector, the optimal weight matrix is $\theta^T \in R^l$, and $l \geq 1$ denotes the number of neurons. The known continuous basis function vector is $\varphi(Z) = [\varphi_1(Z), \varphi_2(Z), \ldots, \varphi_l(Z)]^T$ and $\varphi_i(Z)(i = 1, 2, \ldots,)$ is the Gaussian function with

$$\varphi_i(Z) = \exp\left(-\frac{\|Z - \mu_i\|^2}{\eta_i^2}\right), \quad i = 1, 2, \ldots, l, \tag{10}$$

where $\mu_i = [\mu_1, \ldots, \mu_l]^T, l > 1$ stands for the center of the receptive field, and the width of the Gaussian basis function is η_i.

Remark 1. As usual, the neural network chosen in this article has some special properties. Firstly, with the increase of the number of neurons, approximation error $\varsigma(Z)$ will gradually decrease. Secondly, the approximate area of the NNs is uniformly covered by the receptive field μ_i. Finally, based on the equation $\eta_i = \sqrt{2} d_{\max}/\sqrt{l}$, we can get the width of the Gaussian function and $d_{\max}$ denotes the maximal Euclid distance of the approximation region.

It is necessary to research the marine vessel position constraints because marine environmental disturbances are time-varying and marine conditions are complex. In [27], for the nonlinear system control algorithm based on BLF, the state constraint is converted into the constraint of tracking error, which increases the restriction on the initial conditions of the system. It should be noted that most of the previous research studies on constraints were carried out in simple cases. To deal with time-varying output constraints gracefully, we consider the following nonlinear state-dependent transformation (NSDT):

$$\mu_{1k} = \frac{x_{1k}(t)}{\left(F_{k1}(t) + x_{1k}(t)\right)\left(F_{k2}(t) - x_{1k}(t)\right)}, \tag{11}$$

where the initial states $x_{1k}(0) \in D_{1k}$, in which $k = 1, 2, 3$. From (11), it should be pointed that μ_{1k} tends to infinity when x_{1k} is close to D_{1k}, for any initial condition $x_{1k}(0) \in D_{1k}$. That is

$$\mu_{1k} \longrightarrow \pm\infty \text{ if and only if } x_{1k} \longrightarrow -F_{k1} \text{ or } x_{1k} \longrightarrow F_{k2}. \tag{12}$$

Therefore, it is noted that for any $x_{1k}(0) \in D_{1k}$, if $\mu_{1k} \in L_\infty$, $t \geq 0$, then $x_{1k} \in D_{1k}$ can be guaranteed for any $t \geq 0$, that is only need $\mu_{1k}, k = 1, 2, 3$, are bounded, and time-varying output constraints will not be violated. In other words, the issue of output constraints attributed to making sure that μ_{1k} for $t \geq 0$ is bounded.

Assumption 1. Time-varying function F_{k1}, $k = 1, 2, 3$, and F_{k2}, $k = 1, 2, 3$, are bounded and continuous, as well as its first derivative is also bounded and continuous.

The time derivative of μ_{1k} is given as

$$\dot{\mu}_{1k} = \eta_{1k}\dot{x}_{1k} + \gamma_{1k}, \tag{13}$$

where

$$\begin{aligned} \eta_{1k} &= \frac{F_{k1}F_{k2} + x_{1k}^2}{\left(F_{k1} + x_{1k}\right)^2\left(F_{k2} - x_{1k}\right)^2}, \\ \gamma_{1k} &= \frac{-\left(\dot{F}_{k1}F_{k2} + F_{k1}\dot{F}_{k2}\right)x_{1k} + \left(\dot{F}_{k1} - \dot{F}_{k2}\right)x_{1k}^2}{\left(F_{k1} + x_{1k}\right)^2\left(F_{k2} - x_{1k}\right)^2}, \end{aligned} \tag{14}$$

for $k = 1, 2, 3$.

Then, (13) can also be rewritten as follows:

$$\dot{\mu}_k = \eta_k\dot{x}_k + \gamma_k, \tag{15}$$

where

$$\begin{aligned} \mu_k &= [\mu_{11}, \mu_{12}, \mu_{13}]^T \in R, \quad k = 1, 2, 3, \\ \eta_k &= \text{diag}\{\eta_{1k} \in R^{3\times 3}, \quad k = 1, 2, 3, \\ \gamma_k &= [\gamma_{k1}, \gamma_{k2}, \gamma_{k3}]^T \in R, \quad k = 1, 2, 3. \end{aligned} \tag{16}$$

In this paper, feasibility condition in the control design is removed. Accordingly, it is much better to implement control methods. Most of the existing vessel constraint methods dispose the constant state constraints. For comprehensive consideration, the adaptive NN control of time-varying output constraints is studied in this article.

The different coordinate transformation is introduced as follows:

$$z_1 = \mu_1 - \beta_d, \tag{17}$$

$$z_2 = x_2 - \alpha_1, \tag{18}$$

with

$$\begin{aligned} \beta_d &= [\beta_{d1}, \beta_{d2}, \beta_{d3}]^T \in R, \\ \beta_{dk} &= \frac{\varpi_{dk}(t)}{\left(F_{k1}(t) + \varpi_{dk}(t)\right)\left(F_{k2}(t) - \varpi_{dk}(t)\right)}, \quad k = 1, 2, 3, \end{aligned} \tag{19}$$

where α_1 is the intermediate controller and it will be defined later. It should be clear that β_{dk} is bounded in the set of D_{1k}.

Assumption 2. The expected tracking signal ϖ_{dk}, $k = 1, 2, 3$, is bounded, and the derivative of ϖ_{dk} with respect to time is also bounded.

Then, the time derivative of β_d is given as follows:

$$\dot{\beta}_d = \eta_d\dot{\varpi}_d + \rho_d, \tag{20}$$

where

$$\begin{aligned}
&\eta_d = \mathrm{diag}\{\eta_{dk}\} \in R^{3\times 3}, \quad k = 1,2,3,\\
&\rho_{dk} = [\rho_{d1}, \rho_{d2}, \rho_{d3}]^T \in R, \quad k = 1,2,3,\\
&\varpi_{dk} = [\varpi_{d1}, \varpi_{d2}, \varpi_{d3}]^T \in R, \quad k = 1,2,3,\\
&\eta_{dk} = \frac{F_{k1}F_{k2} + \varpi_{dk}^2}{(F_{k1} + \varpi_{dk})^2 (F_{k2} - \varpi_{dk})^2},\\
&\rho_{dk} = \frac{-(\dot{F}_{k1}F_{k2} + F_{k1}\dot{F}_{k2})\varpi_{dk} + (\dot{F}_{k1} - \dot{F}_{k2})\varpi_{dk}^2}{(F_{k1} + \varpi_{dk})^2 (F_{k2} - \varpi_{dk})^2},
\end{aligned} \tag{21}$$

for $k = 1, 2, 3$, and according to the definition of η_{dk} and ρ_{dk}, we know that η_{dk} and ρ_{dk} are known and computable and will be used in the design of the controller in the later.

Remark 2. For uncertain vessel nonlinear system, to make sure that the state constraints are never violated their bounded, the virtual controller must satisfy the feasibility condition:

$$-F_{k1} \le \alpha_{k1} \le F_{k2}, \quad k = 1,2,3, \tag{22}$$

where α_{k1} is the virtual controller and $-F_{k1}$ and F_{k2} are smooth functions. Although the virtual controller α_{k1} is related to state variables, it also depends on other design parameters. The feasibility conditions of (22) must be satisfied, and only in this way can the control scheme be implemented effectively. However, it is unrealistic and even difficult to find such parameters.

Remark 3. It should be pointed that according to the actual need, the external disturbances are required to be bounded. Note that all of these boundary conditions are not necessary to implement adaptive control, and they can only be used in analysis.

Remark 4. If l and m are the nonnegative real numbers, $a > 1$, b is nonzero real number, and $1/a + 1/b = 1$. Then, we can get

$$lm \le \frac{l^a}{a} + \frac{m^b}{b}. \tag{23}$$

In this article, the following disturbance observer is constructed to deal with unknown disturbance:

$$\begin{aligned}
&\hat{g} = \gamma + K_0 M x_2,\\
&\dot{\gamma} = -K_0\gamma - K_0[-A(x_2)x_2 - B(x_2)x_2 - f(x_1)\\
&\qquad + \tau + K_0 M x_2],
\end{aligned} \tag{24}$$

where $K_0 \in R^{3\times 3}$ is an observer gain positive definite matrix, γ is the introduced intermediate variable, and the definition of $\hat{g}$ is given in the following equation:

$$\begin{aligned}
\dot{\tilde{g}} &= \dot{\gamma} + K_0 M \dot{x}_2\\
&= -K_0\gamma - K_0[-A(x_2)x_2 - B(x_2)x_2 - f(x_1) + \tau\\
&\quad + K_0 M x_2]\\
&\quad + K_0[\tau + g - A(x_2)x_2 - B(x_2)x_2 - f(x_1)]\\
&= K_0[g - (\gamma + K_0 M x_2)] = K_0(g - \hat{g}).
\end{aligned} \tag{25}$$

Then, differentiating $\tilde{g}$, it yields

$$\dot{\tilde{g}} = \dot{g} - \dot{\hat{g}} = \dot{g} - K_0(g - \hat{g}) = \dot{g} - K_0\tilde{g}, \tag{26}$$

where $\tilde{g} = g - \hat{g}$ represents estimation error vector and $\tilde{g} = [\tilde{g}_1, \tilde{g}_2, \tilde{g}_3]^T$, vector $\hat{g} = [\hat{g}_1, \hat{g}_2, \hat{g}_3]^T$ denotes disturbance estimation, and vector $g = [g_1, g_2, g_3]^T$ is external disturbance. Define $\|\dot{g}(t)\| \le C_g \le \infty$, where C_g is a positive constant.

The detailed design steps are as follows:

Step 1: the following Lyapunov function is chosen:

$$V_1 = \frac{1}{2} z_1^T z_1. \tag{27}$$

Based on $z_1 = \mu_1 - \beta_d$ in (17), and the definition of system (8), then taking derivation of z_1, one obtains

$$\begin{aligned}
\dot{z}_1 &= \dot{\mu}_1 - \dot{\beta}_d\\
&= \eta_1 R(x_1)x_2 + \gamma_1 - \eta_d\dot{\varpi}_d - \rho_d\\
&= \eta_1 R(x_1)z_2 + \eta_1 R(x_1)\alpha_1 + \gamma_1 - \eta_d\dot{\varpi}_d - \rho_d.
\end{aligned} \tag{28}$$

Based on (28), the time derivative of V_1 is given by

$$\begin{aligned}
\dot{V}_1 &= z_1^T \dot{z}_1\\
&= z_1^T(\eta_1 R(x_1)z_2 + \eta_1 R(x_1)\alpha_1 + \gamma_1 - \eta_d\dot{\varpi}_d - \rho_d).
\end{aligned} \tag{29}$$

The virtual controller is designed as follows:

$$\alpha_1 = R^T(x_1)\eta_1^{-1}(-K_1 z_1 - \gamma_1 + \eta_d\dot{\varpi}_d + \rho_d). \tag{30}$$

Substituting (30) into (29), we have

$$\dot{V}_1 = -z_1^T K_1 z_1 + z_1^T \eta_1 R(x_1) z_2. \tag{31}$$

Step 2: the Lyapunov function is designed as follows:

$$V_2 = \frac{1}{2} z_2^T M z_2 + \frac{1}{2}\tilde{g}^T\tilde{g} + \frac{1}{2}\sum_{i=1}^{3}\tilde{\theta}_i^T\Gamma_i^{-1}\tilde{\theta}_i, \tag{32}$$

where $\widetilde{\theta} = \theta - \widehat{\theta}$ denotes the NN weight error, $\widehat{\theta}$ is the estimation of weight matrix θ, and $\Gamma_i = \Gamma_i^T$ is positive definite matrix.

Taking the derivative of z_2 in (18), and based on $\dot{x}_2 = M^{-1}(\tau + g - f(x_1) - A(x_2)x_2 - B(x_2)x_2)$, it yields

$$\begin{aligned}\dot{z}_2 &= \dot{x}_2 - \dot{\alpha}_1 \\ &= M^{-1}(\tau + g - f(x_1) - A(x_2)x_2 - B(x_2)x_2) - \dot{\alpha}_1.\end{aligned} \tag{33}$$

Substituting (33) into (32), it leads to

$$\begin{aligned}\dot{V}_2 &= z_2^T M\dot{z}_2 + \widetilde{g}^T\dot{\widetilde{g}} + \sum_{i=1}^{3}\widetilde{\theta}_i^T\Gamma_i^{-1}\dot{\widetilde{\theta}}_i \\ &= z_2^T(\tau + g - f(x_1) - A(x_2)x_2 - B(x_2)x_2 - M\dot{\alpha}_1) \\ &\quad + \widetilde{g}^T\dot{\widetilde{g}} + \sum_{i=1}^{3}\widetilde{\theta}_i^T\Gamma_i^{-1}\dot{\widetilde{\theta}}_i.\end{aligned} \tag{34}$$

The unknown function is defined by

$$F_2(Z_2) = -A(x_2)x_2 - B(x_2)x_2 - M\dot{\alpha}_1 - f(x_1), \tag{35}$$

where $F_2(Z_2)$ is an unknown and continuous function and $Z_2 = [x_1^T, x_2^T, \alpha_1^T, \dot{\alpha}_1^T]$ are the inputs of the NNs. With the help of the NN approximation capability, the unknown function can be rewritten as follows:

$$F_2(Z_2) = \theta^T\varphi(Z) + \varsigma(Z). \tag{36}$$

The actual controller τ is constructed as follows:

$$\tau = -R^T(x_1)\eta_1^T z_1 - K_2 z_2 - \widehat{g} - \widehat{\theta}^T\varphi(Z). \tag{37}$$

Substituting (37) and (36) into (34), one obtains

$$\begin{aligned}\dot{V}_2 \le\ & -z_2^T K_2 z_2 - z_2^T R^T(x_1)\eta_1^T z_1 + z_2^T\widetilde{g} + z_2^T\widetilde{\theta}^T\varphi(Z) \\ & + z_2^T\varsigma(Z) + \widetilde{g}^T\dot{\widetilde{g}} - \sum_{i=1}^{3}\widetilde{\theta}_i^T\Gamma_i^{-1}\dot{\widetilde{\theta}}_i.\end{aligned} \tag{38}$$

Substituting (26) into (38), we have

$$\begin{aligned}\dot{V}_2 &\le -z_2^T K_2 z_2 - z_2^T R^T(x_1)\eta_1^T z_1 + z_2^T\widetilde{g} + z_2^T\varsigma(Z) \\ &\quad + z_2^T\widetilde{\theta}^T\varphi(Z) + \widetilde{g}^T(\dot{g} - K_0\widetilde{g}) - \sum_{i=1}^{3}\widetilde{\theta}_i^T\Gamma_i^{-1}\dot{\widehat{\theta}}_i \\ &\le -z_2^T K_2 z_2 - z_2^T R^T(x_1)\eta_1^T z_1 + z_2^T\widetilde{g} + z_2^T\varsigma(Z) \\ &\quad + z_2^T\widetilde{\theta}^T\varphi(Z) + \widetilde{g}^T\dot{g} - \widetilde{g}^T K_0\widetilde{g} - \sum_{i=1}^{3}\widetilde{\theta}_i^T\Gamma_i^{-1}\dot{\widehat{\theta}}_i.\end{aligned} \tag{39}$$

The adaptive laws are selected as follows:

$$\dot{\widehat{\theta}}_i = \Gamma_i\left(\varphi_i(Z_i)z_{2,i} - \delta_i\widehat{\theta}_i\right), \quad i = 1, 2, 3, \tag{40}$$

where $\delta_i > 0$, $i = 1, 2, 3$ is the design parameter.

Using Young's inequality, we have the following inequalities:

$$\begin{aligned}z_2^T\widetilde{g} &\le h_1 z_2^T z_2 + \frac{1}{4h_1} \\ \widetilde{g}^T\widetilde{g}\widetilde{g}^T\dot{g} &\le h_2\widetilde{g}^T\widetilde{g} + \frac{1}{4h_2}C_g \\ z_2^T\varsigma(Z) &\le \frac{1}{2}z_2^T z_2 + \frac{1}{2}\overline{\varsigma}^2 \\ \sum_{i=1}^{3}\delta_i\widetilde{\theta}_i^T\widehat{\theta}_i &= \sum_{i=1}^{3}\delta_i\widetilde{\theta}_i^T\theta_i - \sum_{i=1}^{3}\delta_i\widetilde{\theta}_i^T\widetilde{\theta}_i \\ &\le -\frac{1}{2}\sum_{i=1}^{3}\delta_i\|\widetilde{\theta}_i\|^2 + \frac{1}{2}\sum_{i=1}^{3}\delta_i\|\theta_i\|^2,\end{aligned} \tag{41}$$

where h_1 and h_2 are the design parameters.

Substituting all the above inequalities into (39), we have

$$\begin{aligned}\dot{V}_2 \le\ & -z_2^T K_2 z_2 - z_2^T R^T(x_1)\eta_1^T z_1 + h_1 z_2^T z_2 + \frac{1}{2}z_2^T z_2 \\ & + \frac{1}{4h_1}\widetilde{g}^T\widetilde{g} + h_2\widetilde{g}^T\widetilde{g} - \widetilde{g}^T K_0\widetilde{g} - \frac{1}{2}\sum_{i=1}^{3}\delta_i\|\widetilde{\theta}_i\|^2 \\ & + \frac{1}{2}\sum_{i=1}^{3}\delta_i\|\theta_i\|^2 + \frac{1}{4h_2}C_g + \frac{1}{2}\overline{\varsigma}^2.\end{aligned} \tag{42}$$

Consider the Lyapunov function as follows:

$$\begin{aligned}V &= V_1 + V_2 \\ &= \frac{1}{2}z_1^T z_1 + \frac{1}{2}z_2^T M z_2 + \frac{1}{2}\widetilde{g}^T\widetilde{g} + \frac{1}{2}\sum_{i=1}^{3}\widetilde{\theta}_i^T\Gamma_i^{-1}\widetilde{\theta}_i.\end{aligned} \tag{43}$$

From (31) and (42), we can obtain

$$\begin{aligned}\dot{V} \le\ & -z_1^T K_1 z_1 + z_1^T\eta_1 R(x_1)z_2 - z_2^T K_2 z_2 - z_2^T R^T(x_1)\eta_1^T z_1 \\ & + h_1 z_2^T z_2 + \frac{1}{2}z_2^T z_2 + \frac{1}{4h_1}\widetilde{g}^T\widetilde{g} + h_2\widetilde{g}^T\widetilde{g} - \widetilde{g}^T K_0\widetilde{g} \\ & - \frac{1}{2}\sum_{i=1}^{3}\delta_i\|\widetilde{\theta}_i\|^2 + \frac{1}{2}\sum_{i=1}^{3}\delta_i\|\theta_i\|^2 + \frac{1}{4h_2}C_g + \frac{1}{2}\overline{\varsigma}^2 \\ \le\ & -z_1^T K_1 z_1 - z_2^T K_2 z_2 + h_1 z_2^T z_2 + \frac{1}{4h_1}\widetilde{g}^T\widetilde{g} \\ & + h_2\widetilde{g}^T\widetilde{g} + \frac{1}{2}z_2^T z_2 - \widetilde{g}^T K_0\widetilde{g} - \frac{1}{2}\sum_{i=1}^{3}\delta_i\|\widetilde{\theta}_i\|^2 \\ & + \frac{1}{2}\sum_{i=1}^{3}\delta_i\|\theta_i\|^2 + \frac{1}{4h_2}C_g + \frac{1}{2}\overline{\varsigma}^2 \le -CV + D,\end{aligned} \tag{44}$$

where

$$C = -2\min\left[\lambda_{\min}(K_1), \lambda_{\min}\left(K_2 M^{-1}\right) - h_1\lambda_{\max}\left(M^{-1}\right) - \frac{1}{2}\lambda_{\max}\left(M^{-1}\right), \lambda_{\min}(K_0) - \frac{1}{4h_1} - h_2, \frac{1}{2}\delta_i\lambda_{\min}(\Gamma_i)\right],$$
$$D = \frac{1}{2}\sum_{i=1}^{3}\delta_i\left\|\theta_i\right\|^2 + \frac{1}{4h_2}C_g + \frac{1}{2}\bar{\zeta}^2, \tag{45}$$

with $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ standing for the largest and minimum eigenvalue of a matrix, respectively.

Theorem 1. *Consider the marine vessel nonlinear system with unknown time-varying disturbances (8) under the virtual controller (30), actual controller (37), and adaptive laws (40). Then, we have the following: (1) error signal undulates in a small neighborhood close to zero; (2) all signals in the closed-loop system are SGUUB; and (3) in the meantime, the output signal will not violate the time-varying boundary.*

Proof. According to (44), we can obtain

$$0 \le V(t) \le \frac{D}{C} + \left(V(0) - \frac{D}{C}\right)e^{-Ct},$$
$$\left\|z_1\right\| \le \sqrt{\frac{2D}{C} + 2\left(V(0) - \frac{D}{C}\right)e^{-Ct}}. \tag{46}$$

From (44), it is clear that the signals z_1, z_2, $\widetilde{\theta}_i$, $i = 1, 2, 3$, and $\widetilde{g}$ are SGUUB. Since $\widehat{\theta}_i$, $i = 1, 2, 3$ is bounded by $\widetilde{\theta}_i = \theta_i - \widehat{\theta}_i$ with θ_i being also bounded. As well, b is bounded. Because β_{dk} is bounded in the set D_{1k}, from $z_1 = \mu_1 - \beta_d$, we can infer that $\mu_{1k} \in L_\infty$, $k = 1, 2, 3$. Based on (12), for any initial condition $x_{1k}(0) \in D_{1k}$, $k = 1, 2, 3$, one deduces that $x_{1k}(t)$ is bounded. Because η_d, ρ_d, $\dot{\bar{\omega}}_d$, and $\bar{\omega}_d$ are bounded, gradually, it is obtained that the virtual controller α_1 is also bounded. According to $z_2 = x_2 - \alpha_1$, we can infer that x_2 is bounded. □

3. Simulation Example

In this section, to illustrate the effectiveness of the proposed method, we conduct the simulation on the model of Cybership II, which is a 1 : 70 scale supply vessel replica setup in a marine control laboratory in the Norwegian University of Science and Technology [28]. The following expected trajectory is considered:

$$\bar{\omega}_{d1} = -0.2 + 0.04\sin(0.5t),$$
$$\bar{\omega}_{d2} = -0.1 + 0.02\cos(0.2t), \tag{47}$$
$$\bar{\omega}_{d3} = 0.002\sin(t),$$

where the matrices are chosen as follows:

$$M = \begin{bmatrix} 20 & 0 & 0 \\ 0 & 23.6 & -0.8 \\ 0 & -0.8 & 2.76 \end{bmatrix},$$
$$A = \begin{bmatrix} 0 & 0 & -23.6\nu_y - 0.8\nu_\psi \\ 0 & 0 & 20\nu_x \\ 23.6\nu_y + 0.8\nu_\psi & -20\nu_x & 0 \end{bmatrix},$$
$$B = \begin{bmatrix} B_1 & B_2 & B_3 \end{bmatrix}^T,$$
$$B_1 = \begin{bmatrix} 0.2565 + 1.5687\left|\nu_x\right| + 3.4357\nu_x^2 \\ 0 \\ 0 \end{bmatrix}^T,$$
$$B_2 = \begin{bmatrix} 0 \\ 0.3578 + 34.5653\left|\nu_y\right| + 5.21\left|\nu_\psi\right| \\ 0.674\left|\nu_y\right| + 3.76\left|\nu_\psi\right| - 0.1079 \end{bmatrix}^T, \tag{48}$$
$$B_3 = \begin{bmatrix} 0 \\ -1.1052 - 5.0437\left|\nu_y\right| - 0.13\left|\nu_\psi\right| \\ 0.674\left|\nu_y\right| + 3.76\left|\nu_\psi\right| - 0.1079 \end{bmatrix}^T.$$

The disturbance vector is given as follows:

$$g(t) = \begin{bmatrix} 1.4 + 2\sin(0.02t) + 1.5\sin(0.1t) \\ -0.9 + 2\sin\left(0.02t - \frac{\pi}{6}\right) + 1.5\sin(0.3t) \\ -\sin\left(0.09t + \frac{\pi}{3}\right) - 4\sin(0.01t) \end{bmatrix}. \tag{49}$$

The initial state variables are selected as $x_1 = [x_{11}, x_{12}, x_{13}]^T = [0.004, 0.019, 0^T]$ and $x_2 = [x_{21}, x_{22}, x_{23}]^T = [0.4, 0.05, 0]^T$. The state variables x_{11}, x_{12}, and x_{13} satisfy $-F_{11} \le x_{11} \le F_{12}$, $-F_{21} \le x_{12} \le F_{22}$, and $-F_{31} \le x_{13} \le F_{32}$ with $F_{11} = 1 + 0.1\sin(t)$, $F_{12} = 1 + 0.1\sin(t)$, $F_{21} = 1 + 0.1\sin(t)$, $F_{22} = 1 + 0.1\sin(t)$, $F_{13} = 0.3 + 0.1\sin(t)$, and $F_{23} = 0.3 + 0.1\sin(t)$.

The initial value of the observer is selected as $\widehat{g}(0) = [0, 0, 0]^T$. The design parameters are given as $\delta = [0.001, 0.001, 0.001]^T$. The parameter matrices are given as $\Gamma_1 = 50I$, $\Gamma_2 = 100I$, $\Gamma_3 = 200I$, and $K_0 = \mathrm{diag}[2, 2, 2]$. The

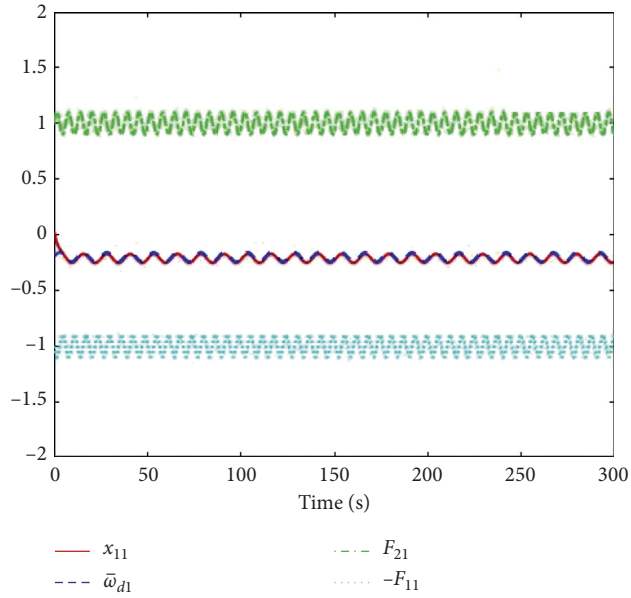


Figure 1: Trajectory of ϖ_{d1}, x_{11}, and constraint intervals.

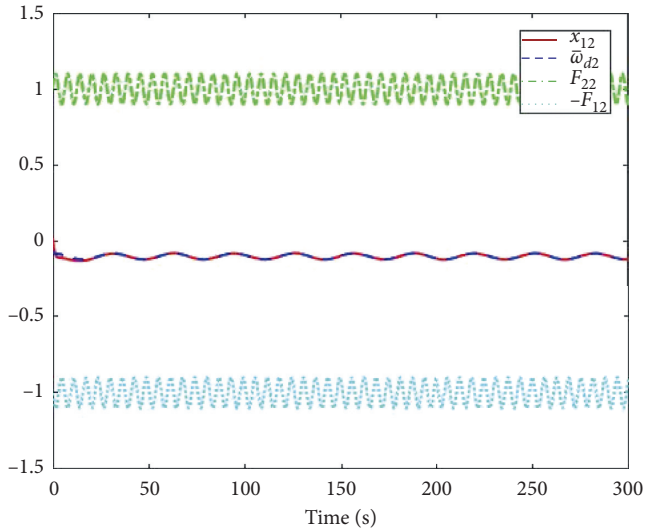


Figure 2: Trajectory of ϖ_{d2}, x_{12}, and constraint intervals.

control gains are set as $K_1 = \text{diag}[0.001, 0.1, 0.1]$ and $K_2 = \text{diag}[200, 200, 0.02]$.

Figures 1–11 show the simulation results. Figures 1–3 show the output signal x_1, reference signal ϖ_{dk}, and constraints interval F_{k1} and F_{k2} with $k = 1, 2, 3$, which indicates that the tracking trajectory is perfect, and the state variable does not violate the constraint interval. Figure 4 indicates the trajectory of the state variable x_2. Figures 5 and 6 are given to illustrate the trajectories of tracking error z_1 and z_2, and it can be seen that the error is small enough. The trajectory of actual controller is displayed in Figure 7. From Figure 8, it indicates that the disturbance observer is constructed to deal with the external disturbance better. Meanwhile, Figures 9 and 10 are used to explain the trajectory of adaptive laws. We can clearly see that the adaptive laws are bounded. Figure 11 shows the trajectory of virtual controller. The results show that the adaptive control method is effective.

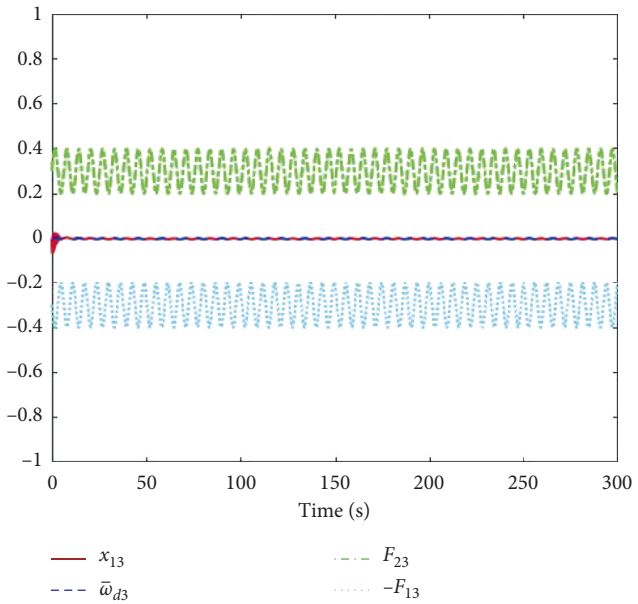


Figure 3: Trajectory of ϖ_{d3}, x_{13}, and constraint intervals.

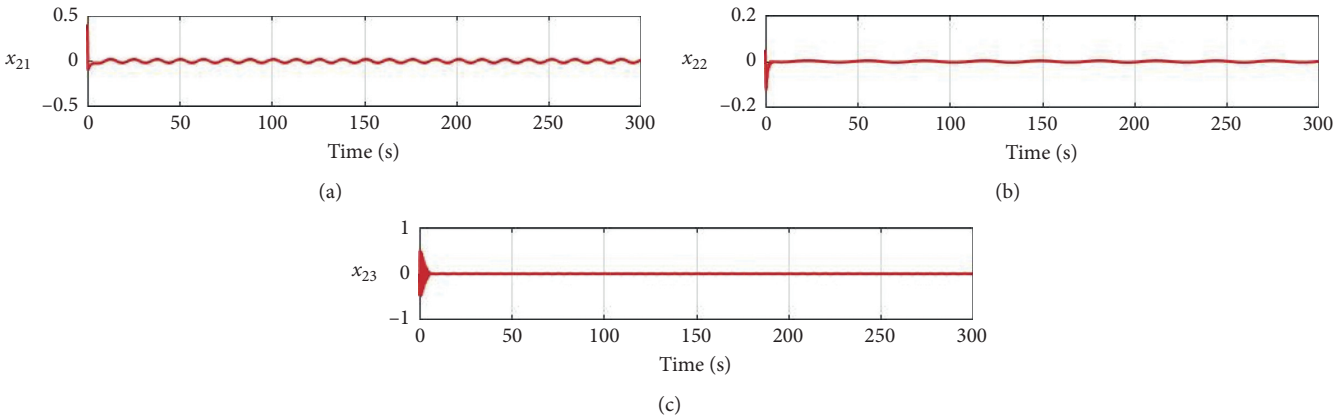


Figure 4: Trajectory of velocity vector x_2.

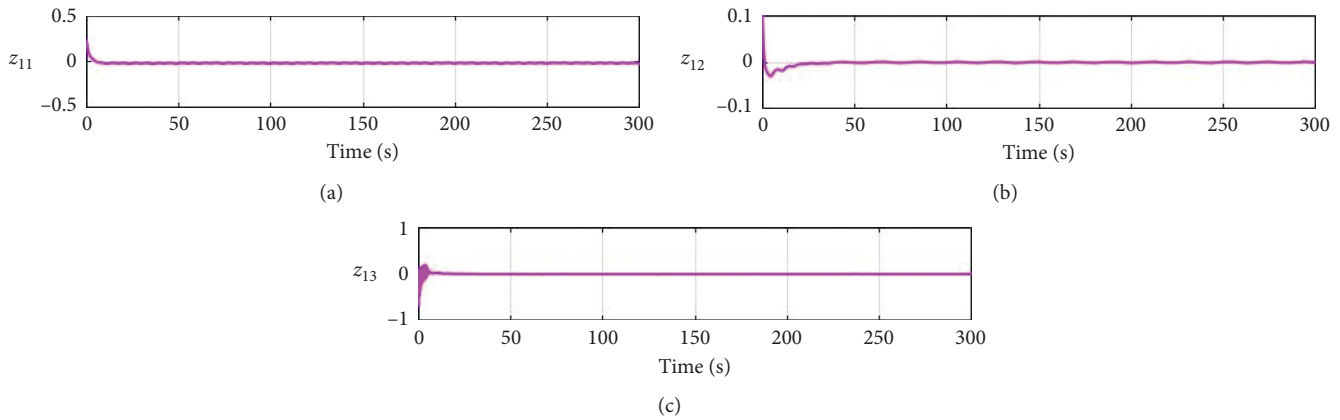


Figure 5: Trajectories of errors z_1.

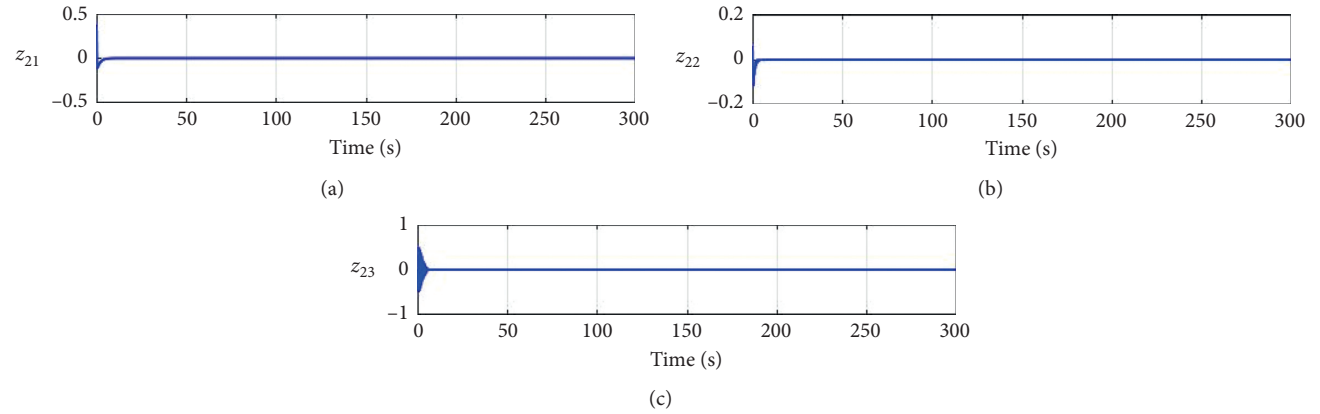


Figure 6: Trajectories of errors z_2.

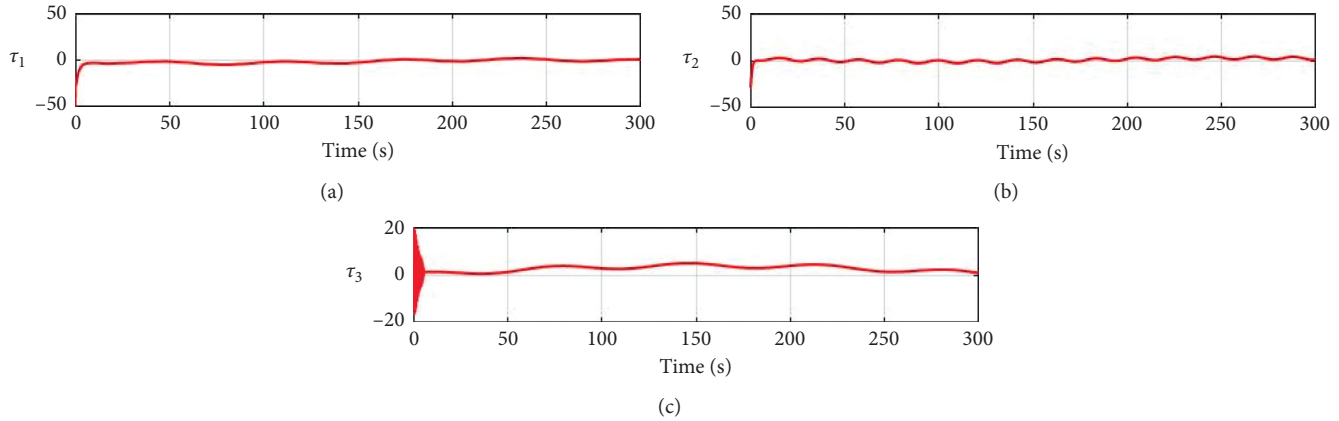


Figure 7: Trajectories of control input τ.

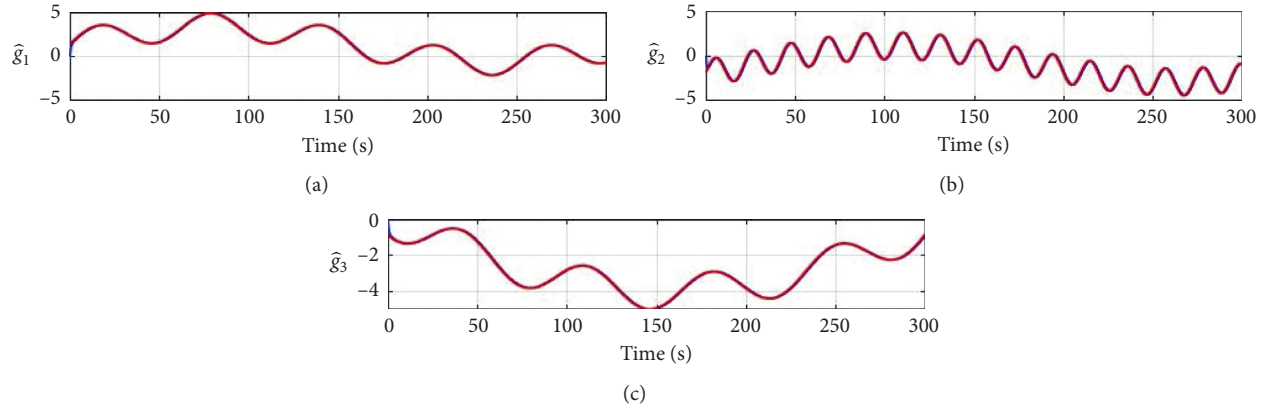


Figure 8: Trajectories of external disturbances g_1, g_2, and g_3 and their estimations $\hat{g}_1$, $\hat{g}_2$, and $\hat{g}_3$.

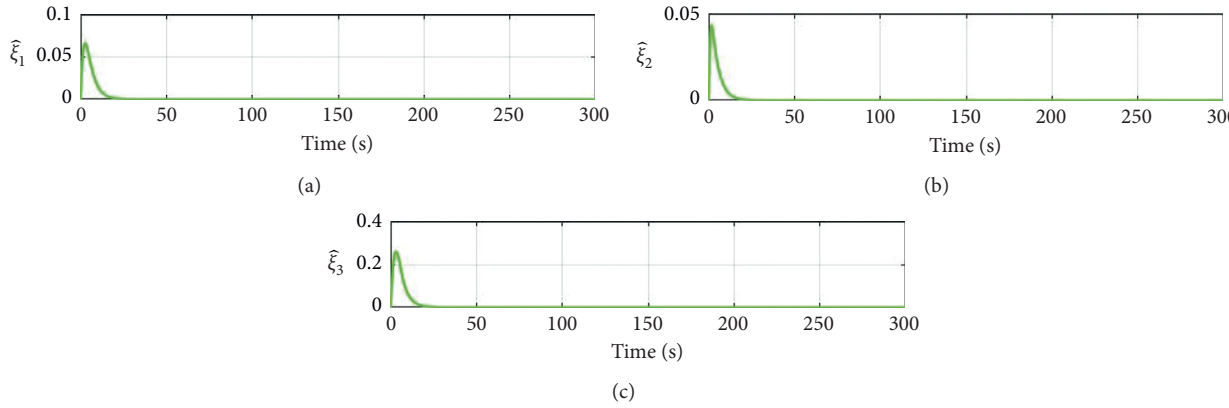


Figure 9: Trajectories of $\hat{\xi}$.

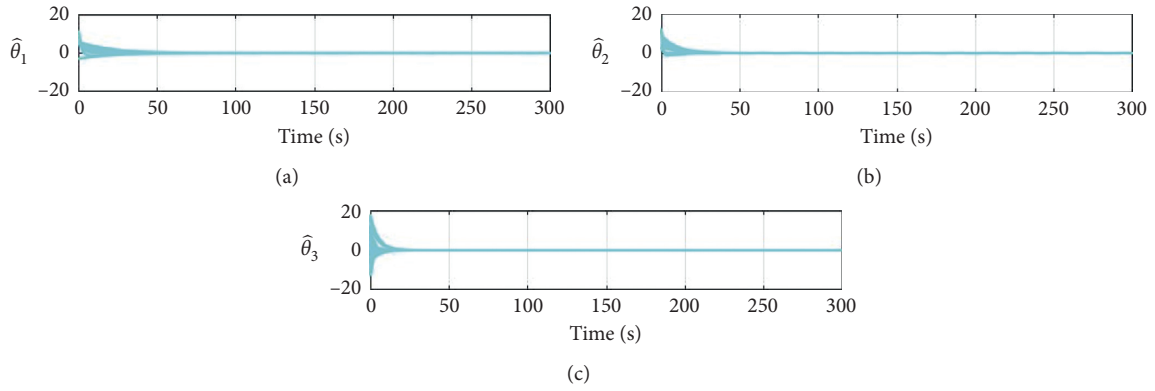


Figure 10: Trajectories of $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_3$.

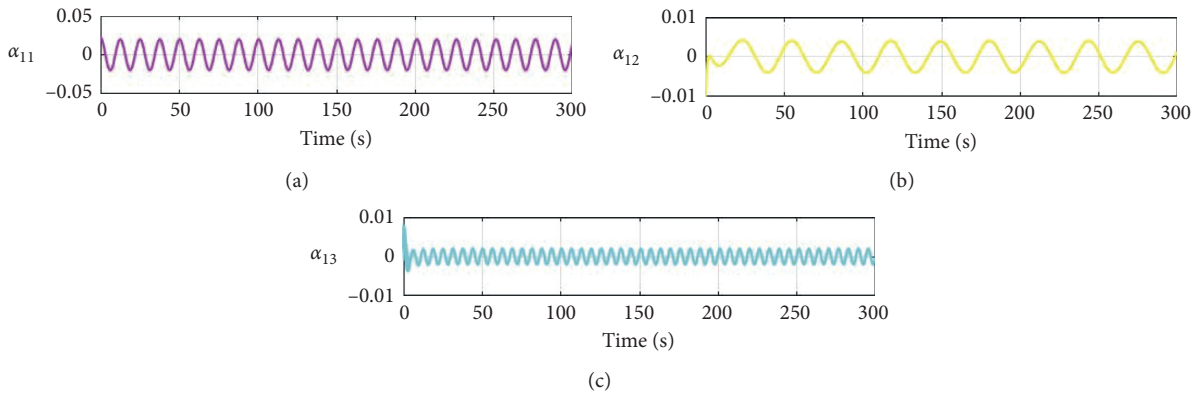


Figure 11: Trajectories of virtual controller α_{11}, α_{12}, and α_{13}.

4. Conclusion

The adaptive NN control method is studied for vessel nonlinear systems with disturbance observer and time-varying output constraints in this paper. Compared with the existing research on vessel, the method studied in this article is more practical. First, the feasibility condition of virtual controller is eliminated by introducing nonlinear state correlation function and new coordinate transformation. Only in this way, we can deal with control design gracefully. Then, with the help of NNs, the unknown and continuous function is approximated to simplify the design of the controller. Next, the disturbance observer is constructed to dispose the disturbances. Finally, the control performance is tested through simulation examples. In the future work, we will apply the constraint form of this paper

to the position and speed constraints of flexible manipulator.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 61973147 and 61903168, in part by the Doctoral Research Initiation of Foundation of Liaoning Province under Grant 2020-BS-243, and in part by the Natural Scientific Research Foundation Guiding Plan of Liaoning Province under Grant 2019-ZD-0701.

References

[1] Y. Xiao, C. L. P. Chen, and B. Wang, "Bandwidth degradation QoS provisioning for adaptive multimedia in wireless/mobile networks," *Computer Communications*, vol. 25, no. 13, pp. 1153–1161, 2002.

[2] L. Cheng, Y. Liu, Z. G. Hou, M. Tan, D. J. Du, and M. R. Fei, "A rapid spiking neural network approach with an application on hand gesture recognition," *IEEE Transactions on Cognitive and Developmental Systems*, 2019.

[3] T. Li, R. Zhao, C. L. P. Chen, L. Fang, and C. Liu, "Finite-time formation control of under-actuated ships using nonlinear sliding mode control," *IEEE Transactions on Cybernetics*, vol. 48, no. 11, pp. 3243–3253, 2018.

[4] T. Li, Z. Li, D. Wang, and C. L. Chen, "Output-feedback adaptive neural control for stochastic nonlinear time-varying delay systems with unknown control directions," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 6, pp. 1188–1201, 2015.

[5] J. Yu, B. Chen, H. Yu, C. Lin, and L. Zhao, "Neural networks-based command filtering control of nonlinear systems with uncertain disturbance," *Information Sciences*, vol. 426, pp. 50–60, 2018.

[6] L.-B. Wu and J. H. Park, "Adaptive fault-tolerant control of uncertain switched nonaffine nonlinear systems with actuator faults and time delays," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 9, pp. 3470–3480, 2020.

[7] L. Tang, D. Ma, and J. Zhao, "Adaptive neural control for switched non-linear systems with multiple tracking error constraints," *IET Signal Processing*, vol. 13, no. 3, pp. 330–337, 2019.

[8] S. Tong, X. Min, and Y. Li, "Observer-based adaptive fuzzy tracking control for strict-feedback nonlinear systems with unknown control gain functions," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3903–3913, 2020.

[9] L. Liu, Y.-J. Liu, D. Li, S. Tong, and Z. Wang, "Barrier Lyapunov function-based adaptive fuzzy FTC for switched systems and its applications to resistance-inductance-capacitance circuit system," *IEEE Transactions on Cybernetics*, vol. 50, no. 8, pp. 3491–3502, 2020.

[10] Q. Cui, Y. Wang, and Y. Song, "Neuroadaptive fault-tolerant control under multiple objective constraints with applications to tire production systems," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2020.

[11] Z. Wang, P. Shi, and C.-C. Lim, "H−/H∞ fault detection observer in finite frequency domain for linear parameter-varying descriptor systems," *Automatica*, vol. 86, pp. 38–45, 2017.

[12] S. Zhang, D. Huang, and W. He, "Active vibration control for a flexible string system with input backlash," *IET Control Theory & Applications*, vol. 10, no. 7, pp. 800–805, 2016.

[13] H. Liu and Y. Huang, "Robust adaptive output feedback tracking control for flexible-joint robot manipulators based on singularly perturbed decoupling," *Robotica*, vol. 36, no. 6, pp. 822–838, 2018.

[14] W. He and S. S. Ge, "Cooperative control of a nonuniform gantry crane with constrained tension," *Automatica*, vol. 66, no. 4, pp. 146–154, 2016.

[15] Z. Liu, J. Liu, and W. He, "Modeling and vibration control of a flexible aerial refueling hose with variable lengths and input constraint," *Automatica*, vol. 77, pp. 302–310, 2017.

[16] K. P. Tee, S. S. Ge, and E. H. Tay, "Barrier Lyapunov functions for the control of output-constrained nonlinear systems," *Automatica*, vol. 45, no. 4, pp. 918–927, 2009.

[17] D.-P. Li, Y.-J. Liu, S. Tong, C. L. P. Chen, and D.-J. Li, "Neural networks-based adaptive control for nonlinear state constrained systems with input delay," *IEEE Transactions on Cybernetics*, vol. 49, no. 4, pp. 1249–1258, 2019.

[18] K. P. Tee and S. S. Ge, "Control of state-constrained nonlinear systems using integral barrier Lyapunov functionals," in *Proceedings of the 2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pp. 3239–3244, Maui, HI, USA, December 2012.

[19] L. Liu, Y. J. Liu, S. C. Tong, and C. L. P. Chen, "Integral barrier Lyapunov function based adaptive control for switched nonlinear systems," *Science China Information Sciences*, vol. 63, no. 3, Article ID 132203, 2020.

[20] L. Liu and X. Li, "Event-triggered tracking control for active seat suspension systems with time-varying full-state constraints," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, p. 1, 2020.

[21] D. Li, C. L. P. Chen, Y.-J. Liu, and S. Tong, "Neural network controller design for a class of nonlinear delayed systems with time-varying full-state constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2625–2636, 2019.

[22] Y.-J. Liu, L. Ma, L. Liu, S. Tong, and C. L. P. Chen, "Adaptive neural network learning controller design for a class of nonlinear systems with time-varying state constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 1, pp. 66–75, 2020.

[23] Z.-P. Jiang, "Global tracking control of underactuated ships by Lyapunov's direct method," *Automatica*, vol. 38, no. 2, pp. 301–309, 2002.

[24] N. Khaled and N. Chalhoub, "A dynamic model and a robust controller for a fully-actuated marine surface vessel," *Journal of Vibration and Control*, vol. 17, no. 6, pp. 801–812, 2011.

[25] S.-L. Dai, C. Wang, and F. Luo, "Identification and learning control of ocean surface ship using neural networks," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 4, pp. 801–810, 2012.

[26] S. Yin and B. Xiao, "Tracking control of surface ships with disturbance and uncertainties rejection capability," *IEEE/*

ASME Transactions on Mechatronics, vol. 22, no. 3, pp. 1154–1162, 2016.

[27] K. Zhao, Y. D. Song, T. D. Ma, and L. He, "Prescribed performance control of uncertain Euler-Lagrange systems subject to full state constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3478–3489, 2017.

[28] K. P. Tee and S. S. Ge, "Control of fully actuated ocean surface vessels using a class of feedforward approximators," *IEEE Transactions on Control Systems Technology*, vol. 14, no. 4, pp. 750–756, 2006.