University of Wisconsin Milwaukee
UWM Digital Commons

Theses and Dissertations

August 2015

Intelligent Novel Methods for Identifying Critical Components and Their Combinations for Hypothesized Cyber-physical Attacks Against Electric Power Grids

Ming Wang
University of Wisconsin-Milwaukee

INTELLIGENT NOVEL METHODS FOR IDENTIFYING CRITICAL COMPONENTS AND THEIR COMBINATIONS FOR HYPOTHESIZED CYBER-PHYSICAL ATTACKS AGAINST ELECTRIC POWER GRIDS

by

Ming Wang

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Engineering

at

The University of Wisconsin-Milwaukee

August 2015

ABSTRACT
INTELLIGENT NOVEL METHODS FOR IDENTIFYING CRITICAL COMPONENTS AND THEIR COMBINATIONS FOR HYPOTHESIZED CYBER-PHYSICAL ATTACKS AGAINST ELECTRIC POWER GRIDS

by

Ming Wang

The University of Wisconsin-Milwaukee, 2015
Under the Supervision of Dr. Lingfeng Wang

As a revolutionary change to the traditional power grid, the smart grid is expected to introduce a myriad of noteworthy benefits by integrating the advanced information and communication technologies in terms of system costs, reliability, environmental impacts, operational flexibility, etc. However, the wider deployment of cyber networks in the power grid will bring about important issues on power system cyber security. Meanwhile, the power grid is becoming more vulnerable to various physical attacks due to vandalism and probable terrorist attacks. In an envisioned smart grid environment, attackers have more entry points to various parts of the power grid for launching a well-planned and highly destructive attack in a coordinated manner. Thus, it is important to address the smart grid cyber-physical security issues in order to strengthen the robustness and resiliency of the smart grid in the face of various adverse events. One key step of this research topic is to efficiently identify the vulnerable parts of the smart grid.

In this thesis, from the perspective of smart grid cyber-physical security, three critical component combination identification methods are proposed to reveal the potential vulnerability of the smart grid. First, two performance indices based critical component combination recognition methods are proposed for more effectively identifying the critical component combinations in the multi-component attack scenarios. The optimal selection of critical components is determined according to the criticality of the components, which can be modeled by various performance indices. Further, the space-pruning based enumerative search strategy is investigated to comprehensively and effectively identify critical combinations of multiple same or different types of components. The pruned search space is generated based on the criticality of potential target component which is obtained from low-order enumeration data. Specifically, the combinatorial line-generator attack strategy is investigated by exploring the strategy for attacking multiple different types of components. Finally, an effective, novel approach is proposed for identifying critical component combinations, which is termed search space conversion and reduction strategy based intelligent search method (SCRIS). The conversion and reduction of the search space is achieved based on the criticality of the components which is obtained from an efficient sampling method. The classic intelligent search algorithm, Particle Swarm Optimization (PSO), is improved and deployed for more effectively identifying critical component combinations.

MATLAB is used as the simulation platform in this study. The IEEE 30, 39, 118 and Polish 2383-bus systems are adopted for verifying the effectiveness of the proposed attack strategies. According to the simulation results, the proposed attack strategies turn out to be effective and computationally efficient. This thesis can provide some useful insight into vulnerability identification in a smart grid environment, and defensive strategies can be developed in view of this work to prevent malicious coordinated multi-component attacks which may initiate cascading failures in a cyber-physical environment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Foremost, I would like to express my deepest gratitude to my advisor Dr. Lingfeng Wang for his continuous support of my study and research, for his patience, motivation, immense knowledge and profound insight in the research field. The thesis will not be possible without his professional mentorship at all stages of this research. Dr. Wang's spirit of being diligent, rigorous and creative in scientific research will have a profound impact on my future career development. I could not imagine having a better advisor and mentor for my graduate study.

I would also like to express my sincere gratitude to Dr. David C. Yu for his help and encouragement during my application and the subsequent study at the University of Wisconsin-Milwaukee.

Besides my advisor, I would like to thank the other members of my thesis committee: Dr. David C. Yu and Dr. Yue Liu for their insightful comments and encouragement. Thank you for your time and effort for serving on the committee from your busy schedules.

Additionally, I want to thank my labmates, Yingmeng Xiang and Jun Tan, for their useful discussions in my research and great help in my life. In particular, I am grateful to Yingmeng Xiang for giving me the first glance of this graduate research.

Last but not the least, I would like to thank my family: my parents, Shengcai Huang and Kangru Wang, as well as my brother and sister-in-law for their unconditional love and support.

Chapter 1 Introduction

1.1 Research Background

Power grids play an increasingly important role with the rapid development of modern society. After decades of evolution, the highly interconnected, complex power systems have been built in most countries. The smart grid is an overview of the next generation of power system, which is characterized by the broader use of communication and information technologies in power generation, transmission, distribution, and consumption. After entering the 21st century, with the rapid development of electronic technologies, the latest cyber and communication equipment and devices are being more widely integrated into the conventional power system to build the smart grid. Even though the smart grid is expected to be more economical, efficient, environmentally friendly as compared with the traditional power grid, new cybersecurity issues along with the physical vulnerabilities are bringing new kinds of risks to the power grid nowadays.

Usually, to ensure the power supply reliability of the power grid, the *N-1* or even *N-1-1*, *N-2* criterion is applied. For a power grid which is in compliance with the *N-1* criterion, if one component fails, in general no further losses could be caused. If multiple components are compromised, the power grid is exposed to danger of cascading failures. Despite the fact that many preventive measures have been enforced, large-scale power outages occur at times. There are different causes which could result in cascading failures, such as defects of power system control software, overloading of transmission lines, failures of critical components, natural calamities and man-made sabotages, and so on. For example, in 2003, the well-known Northeast blackout which was partially due to a software bug in the alarm system and has resulted in billions of economic losses and affected 55 million people [1]. In 2005 in China, the whole Hainan province suffered a blackout due to the powerful typhoon “DAMREY” [2]. In 2012, the India blackout which affected 6.7 billion people was caused by the high load demand and the failure of an EHV substation [3]. On February 25th, 2015, 80% citizens in Pakistan suffered a blackout caused by the loss of one key transmission line as the result of a militant attack [4]. Of all the reasons that can cause cascading failures, man-made sabotages are attracting more and more attention since the terrorist threat to the power grid has been escalating. For the traditional power grid, physical attack is the only feasible way for attackers to disrupt the power grid, however it is difficult to attack multiple parts of the power system at the same time. But in cyber attacks, attackers could be able to compromise multiple components simultaneously through executing elaborate attack plans. Also, the combination of cyber attacks and physical attacks is also a potential threat to the grid, which may cause even more severe disruptions. These attacks may lead to an interruption of local or regional area power supply; in serious cases, cascading failures could be caused. Here, some representative attack-related cases are discussed.

Several cascading failures caused by physical attacks have been reported. For example, in April 2013 in California, a Silicon Valley power substation which supplies electric power to thousands of customers was destroyed by a meticulously planned attack, and it took the utility company almost one month to repair and restore it. This incident served as a warning sign to the power companies and regulators. As a result, new guidelines for physical security are being implemented at the request of the FERC, including identifying vulnerable critical components and implementing security enhancement plans [5].

During the past years, a number of cyber-related accidents have occurred in different power systems, which have raised concerns for the cybersecurity of the power grid in a smart grid environment.

For example, a devastating virus named “Stuxnet” was first discovered in June 2010. It was developed to interfere with the normal operation of the Siemens industrial software which is deemed a severe potential threat to the proper operation of the power grid [6].

As concluded by the Repository for Industrial Security Incidents (RISI)’s 2011 annual report, about 35 percent of the industrial control system (ICS) safety incidents were triggered remotely via the cyber network [7], [8]. Since the ICS and SCADA systems play a key role in ensuring the normal operation of the smart grid, more attention should be paid to the emerging cybersecurity issues.

On November 4, 2006, more than 15 million people in Europe suffered a blackout which was primarily caused by human errors. Deficient communication systems played an important role in initiate and propagate this severe accident [9]. Thus, there is also personnel vulnerability in operating the power grid.

Considering all these potential threats to the power system, in order to strengthen the power grid to resist or sustain such risks, one possible way is to study the mechanism of cascading failures and to identify the weak links of the power grid in order to protect and harden it. An effective way to study the cyber-physical-personnel vulnerability of the power grid is to investigate the probable attack methods against the power grid. In this thesis, the term “attacks” refers to all potential cyber, physical or hybrid cyber-physical attacks in the contemporary power grids. Corresponding defensive schemes could be stipulated accordingly based on the outcomes from this study. For instance, the real-time statuses of critical transmission lines could be continuously monitored by the advanced smart grid equipment, and important power stations and substations should be regularly inspected by adequate field personnel.

Some attack strategies against power grids have been proposed in the existing literature. For example, in [10], a flow betweenness based attack strategy against power grid was proposed. A new critical line identification method is proposed according to the fault chain theory in [11]. In [12], critical line identification methods are proposed based on the complex network theory. In [13], a new method for identifying the most critical substation in the power grid is proposed. In [14], joint line-substation attack model is investigated and a component interdependency graph based attack strategy is proposed.

In recent years, the cyber-physical security issue of smart grid also has attracted much attention. In [15], vulnerability detection and defensive strategies are proposed for the cyber security of smart grid. A novel power control system communication network security assessment method is proposed in [16]. In [17], a cyber-physical vulnerability assessment framework is proposed for the security state evaluation in a smart grid environment. In [18], a cyber-physical security evaluation approach is introduced for the purpose of assessing potential cyber-physical contingencies. In [19], a communication protection system is developed to enhance the reliability of smart grid.

In order to effectively investigate the power system security issues, a powerful cascading failure simulation tool is much needed. For a long time, modeling the cascading outages in power systems was a demanding task. In section 1.2, the development trend of cascading failure models as well as the deployed cascading failure simulation platform are discussed in detail.

1.2 Cascading Failures in the Power Grid

The power system is one of the most complex man-made networks. A typical power system mainly consists of generation, transmission and distribution parts. Electric power flows from generators to distribution substations through transmission lines. Represented by graph theory, substations can be denoted by nodes and transmission lines can be denoted by links [20]. In the present cascading failure simulation platforms, cascading failures usually occur due to the massive power flow shift especially in a high-voltage power grid. In the current cascading failure simulation model, the developmental process of cascading failures can be described in the following ways: The fault threshold of each transmission line in the power grid is determined by its physical properties. Kirchhoff's law states that if one component fails in the power grid, the load of this component will be transferred to its nearby components. In general no further damage would be caused since most modern power grids meet the *N-1* contingency criterion. However, if two or more components are tripped, the power flow carried by these components will be transferred to nearby components. If the nearby lines are pushed beyond their line ratings, these lines will be tripped as well, and more lines would be pushed into an overload condition. In this way, cascading failures may propagate in the power grid. The cascading failure will stop if all the transmission lines are tripped or no line is overloaded. In the worst condition, the entire power grid will collapse due to the cascading failure.

So far, several power system cascading failure models have been proposed to investigate the mechanism of cascading failures, some of the representative models are the OPA model [21], the improved OPA model [22], the hidden failure model [23], and so on. Some representative cascading failure simulation platforms are summarized and their modeling features are listed in Table 1-1 [24]. These platforms are able to simulate the development of cascading failures to some degree. Each platform has its own characteristics, for example, the OPA model is able to reveal the influence of slow dynamics due to the upgrade of the power grid [25], and the hidden failure model can be used to simulate the hidden failures of relay protection systems in the power grid [26]. The TRELSS cascading model has been adopted by the industry to deal with cascading failures because it contains detailed built-in modules for studying AC power flow, voltage collapse, approximate protection functions, and operator actions [27].

Table 1-1 Major features of different simulation platforms

Platform Function	CFS	MATCASC	Manchester	OPA	Hidden Failure	TRELSS	CMU	PSA
Overload	√	√	√	√	√	√	√	√
Islanding	√	√	√	√		√	√	
Under frequency load shed			√			√		
Generator redispatch	√		√	√	√	√	√	√
Generator trip	√					√		
Operator response			√			√		√
Blackout time intervals and repair								√
AC network			√		√	√	√	
Voltage collapse			√			√		
Protection Group						√		
Transient stability			√					
Hidden failure					√			
Load increase and grid upgrade				√				

In this study, all the cascading failure simulations are modeled and conducted using a MATLAB based cascading failure simulator (CFS) [28], which is an open source software tool. The main feature of this cascading failure simulation tool is that every stage of the cascading failure can be simulated and the loss caused by the cascading failure can be evaluated. The improved architecture of the CFS is shown in Figure 1-1. This software is mainly composed of four parts [28]:

1) Performing power flow calculation based on DC power flow calculation model and obtaining power flow of each transmission line;

2) Selecting the target components based on the proposed multiple component attack model;

3) Running the main program of cascading failure simulation and

4) Assessing the damage caused by cascading failures after cascading failures subside in the power grid.

For multi-component attacks, at the beginning of the cascading failure simulation, target components will be disconnected from the original power grid and the updated power grid will be formed, then the DC power flow will be conducted for the updated power grid to obtain the power flow of each transmission line. The power grid will be inspected to see if there are any islands generated at each step of the cascading failure simulation process once the transmission lines are tripped from the grid. In certain instances, new islands will be generated because the transmission lines which connect the main power grid and the islands are disconnected. When the new islands are detected, generators and loads will be dispatched accordingly [28]:

1) Initially, the generators in each island will be ramped up or down trying to meet the load demand. The effect of the dispatch scheme is limited by the capacity of the generating units, the ramp rate and ramp time.

2) Secondly, after the dispatching process in each island is finished, if the capacity of generators are still greater than the demand of loads, generators will be tripped in the following order: the smallest generator will be tripped first until the generated power is less than the load demand in the island, then the loads of each load substation will be shed. The curtailed load amount of each load substation is calculated as follows:

$$\Delta P_{d_j}^i = P_{d_j}^i \left(1 - \frac{\sum_{k=1}^{m} P_{g_k}^i}{\sum_{j=1}^{n} P_{d_j}^i}\right) \tag{2.1}$$

where $\Delta P_{d_j}^i$ is the curtailed load of the j_{th} load bus in i_{th} island, $P_{g_k}^i$ is the capacity of the k_{th} generator and $P_{d_j}^i$ is the load demand of the j_{th} load bus in the i_{th} island, m is the number of generators and n is the number of substations in the i_{th} island.

3) Thirdly, after reaching the supply-demand balance in each subgrid, the DC load flow calculation will be executed to obtain the power flow and each transmission line will be checked to see if it is overloaded.

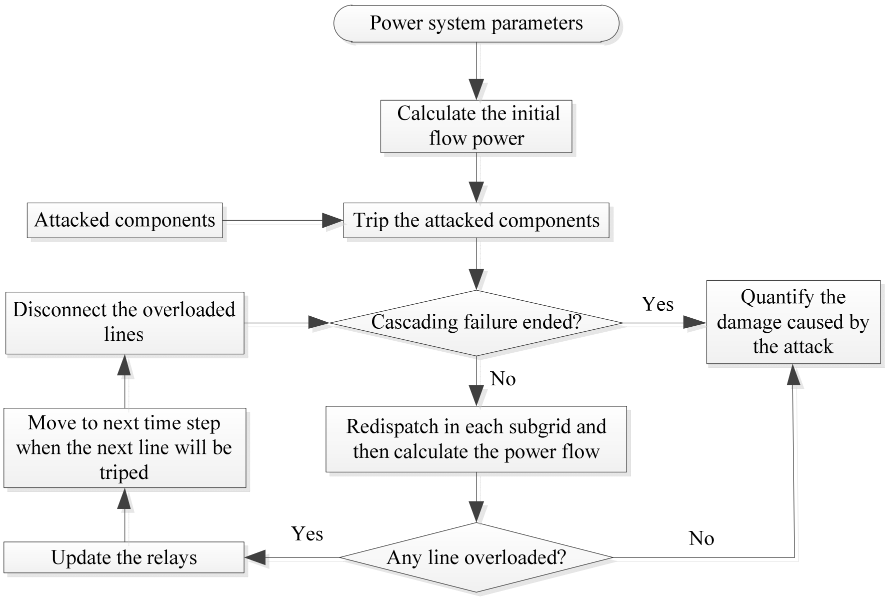

Power system parameters
Calculate the initial flow power
Attacked components
Trip the attacked components
Disconnect the overloaded lines
Cascading failure ended?
Yes
Quantify the damage caused by the attack
No
Move to next time step when the next line will be triped
Redispatch in each subgrid and then calculate the power flow
Yes
No
Update the relays
Any line overloaded?

Figure 1-1 Flow chart of the cascading failure simulator (CFS)

In the cascading failure simulator, overloaded transmission lines are tripped by time delay overcurrent relay. The time delay overcurrent relay considers various factors which could lead to transmission line failures such as overheating of lines and human errors [28].

It is assumed that the power flow of line *i* is P_i and the power flow limit of line *i* is P_i^M. During the cascading failure simulation, the transmission line will be tripped as soon as the accumulated heat exceeds the threshold value T_i. In the simulator, the value of T_i is set to allow the transmission line to continue to run for 5 seconds in the case of overloading 50%. When line *i* is overloaded, the tolerable overloading time is calculated as follows [29]:

$$\Delta_t^i = \begin{cases} \frac{T_i}{P_i - P_i^M} & P_i > P_i^M \\ 0 & otherwise \end{cases} \tag{2.2}$$

Once a transmission line is tripped, the power grid will be updated to see if there are any new islands or overloaded transmission lines until the end of the cascading failure.

After the cascading failure subsides in each island, the consequences of cascading failure will be assessed. Here the index “load loss percentage (P_L)” is adopted. The definition of P_L is shown as follows [30]:

$$P_L = \frac{P_{d0} - P_d}{P_{d0}} \tag{2.3}$$

where P_{d0} is the sum of initial load demand and P_d is the quantity of the survived load demand. The bigger the load loss percentage is, the more serious the consequences caused by the cascading failure are.

In this thesis, some assumptions are made and several definitions are given. The components in the target collections are considered to be tripped from the power grid simultaneously. If an attack scheme can make P_L exceed an expected value, such an attack scheme will be considered as an effective scheme, and the corresponding component combination is regarded as a critical component combination. The expected value is decided by the expectation of the attackers. For instance, if the attackers want to cause the size of a black out as much as 50%, then the attack schemes which can cause P_L higher than 0.5 will be seen as effective scenarios.

1.3 Research Objective and Thesis Layout

The main goal of this thesis is to develop intelligent novel methods for identifying critical components and their combinations to enhance the security and reliability of smart grid.

Throughout this thesis we will examine cascading failures caused by critical combinations of multiple same types and different types of component outages initiated by either potential cyber, physical or hybrid cyber-physical attacks in the contemporary power grids.

In this thesis, three critical component combination identification methods are proposed. In chapter 2, two performance index based critical component combination identification methods are proposed. In chapter 3, the space-pruning enumerative search strategy is investigated to comprehensively and effectively identify critical combinations of multiple same types and different types of components. Further, a more effective approach, search space conversion and reduction strategy based intelligent search method, is introduced in chapter 4. The conclusion are presented and the future work are prospected in chapter 5.

Chapter 2 Critical Component Combination Identification Methods based on Performance Index

2.1 Introduction

In modern power systems, the *N-1* criterion is a basic requirement, which means it is not possible to cause cascading failures if a single component is compromised. However, due to the openness and complexity of the envisioned smart grid, it is possible for intelligent attackers to carry out well-coordinated cyber, physical, or cyber-physical attacks to initiate cascading failures in order to cause large-scale blackouts. For most attackers, they usually have limited attack resources; it is natural for them to select the weakest parts in a power grid as targets in order to achieve the most damaging outcomes with the minimum cost.

From an attacker's perspective, an intuitive way is to select targets based on the criticality of individual components. The central idea of this strategy is to sort the components according to the performance index, then the components with the greatest criticality is chosen as the targets. Several critical component identification methods have been proposed, in [31] a critical node identification method based on the centrality approach is proposed. In [32], a flow transferring index based critical line identification method is suggested.

This chapter investigates the risk of cascading outages caused by simultaneous multi-component attacks. At this stage, for each attack only combinations of multiple same type of components will be considered, like multiple line attacks, multiple substation attacks, etc. Critical component combination identification methods based on performance index will be proposed.

The remainder of this chapter is organized in the following way. The proposed multi-component attack strategies are presented in section 2.2. Case studies and results are presented in section 2.3. The summary of this chapter is given in section 2.4.

2.2 Multi-Component Attack Strategies

When allowing for man-made attacks against power grids, if the attackers only have limited resources to destroy at most *N* components, it is nature for them to try to find out the *N* most critical components based on certain strategies and then damage them to cause as much loss as possible. To strengthen the power grid protection, these identified critical components should be well protected. Therefore, it is urgent to develop an effective identification method to find critical components. In this section, two critical component combination recognition methods, namely static strategy and dynamic strategy, are presented. These strategies are described in detail below.

2.2.1 Static Strategy

The flow chart of the static strategy is given in Figure 2-1. In general, there are four steps to identify the top *N* critical components using static strategy. First, a criticality index should be determined. Second, the criticality index of each component should be calculated based on the selected criticality index. Next, all the components are assumed to be ranked in descending order given by the values of criticality index. Finally, the top N components of the sorted component sequences are chosen as the N most critical components.

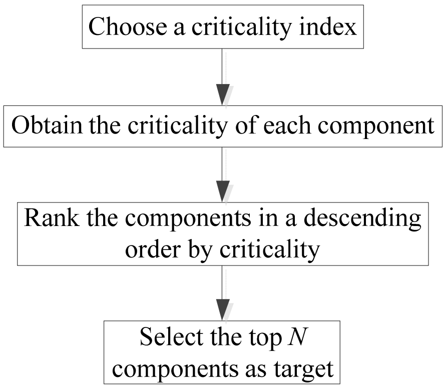

Choose a criticality index
Obtain the criticality of each component
Rank the components in a descending order by criticality
Select the top *N* components as target

Figure 2-1 Flow chart of the static strategy

The criticality index plays an important role when adopting the static strategy. Thus far, many criticality indexes have been proposed, like the edge betweenness centrality [33], the flow betweenness [10], electrical node significance [34], [35], line outage distribution factors [36], and so on. In this study, two criticality indexes, namely the edge betweenness centrality and electrical node significance are implemented.

The edge betweenness centrality is able to find out the greatest central links in an undirected graph. Regarding the topology of the power grid as an undirected graph, a transmission line can be represented by an edge and a substation can be represented by a node, then the edge betweenness centrality E_{BC} of line l can be described as [37]:

$$E_{BC}(l) = \sum_{\substack{i=1 \\ j=1 \\ i \neq j}}^{N_l} \frac{S_{ij}(l)}{S_{ij}} \tag{2.4}$$

where S_{ij} is the number of the shortest routes between node i and node j, and $S_{ij}(l)$ denotes the total sum of the shortest routes roving through line l, N_l is the total number of transmission lines.

The function of electrical node significance index is identifying the node with the largest power flows. In a power grid, the electrical node significance E_N of node j is defined as follows [38]:

$$E_N(i) = \frac{P_i}{\sum_{j=1}^{N} P_j} \tag{2.5}$$

where P_i is the total amount of power flow transferred through node i, and N is the sum of nodes in the power system. To find out the critical nodes based on the index of electrical node significance, the nodes with the greatest electrical node significance will be selected.

To find out critical lines based on the index of electrical node significance, the node with the maximum electrical node significance will be selected first, then the transmission line with the highest load rate will be chosen as the critical line. Power flow is taken into consideration in this method, and the line criticality evaluation is built based on the electrical node significance of each node.

2.2.2 Dynamic Strategy

The flow chart of the dynamic strategy is given in Figure 2-2 and N iterations of selection processes are needed to identify N critical components. Firstly, a criticality index should be determined. In the first iteration, components are rated based on the selected criticality index, then the first target is the component with the greatest criticality index value. The topology of the power grid will be updated considering the first critical component has been removed. Then, in the updated grid, the components will be ranked according to the criticality index and the component with the greatest index value will be selected as the second critical component. The process will be repeated for *N* times to get all the *N* critical components.

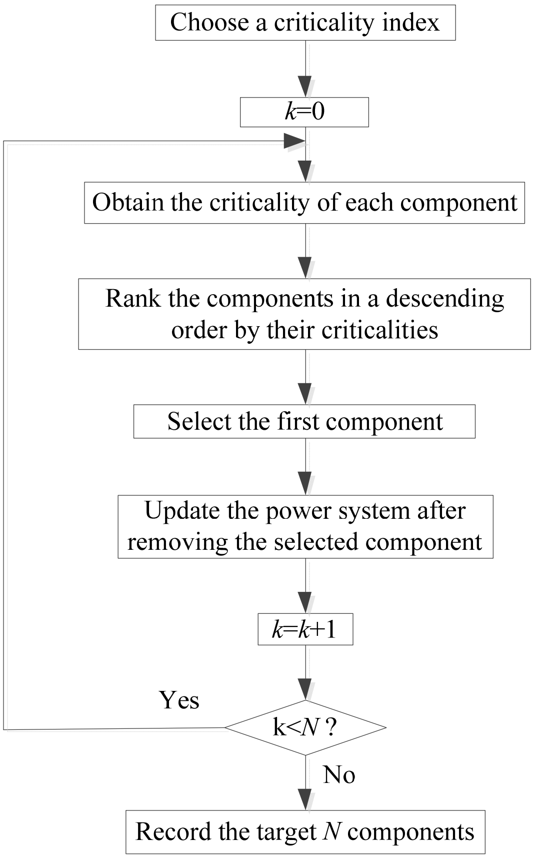

Choose a criticality index
k=0
Obtain the criticality of each component
Rank the components in a descending order by their criticalities
Select the first component
Update the power system after removing the selected component
k=k+1
Yes
k<N ?
No
Record the target N components

Figure 2-2 Flow chart of the dynamic strategy

2.3 Simulations and Results

The IEEE 30-bus test system, IEEE 39-bus test system and IEEE 118-bus test system are adopted here for the cascading failure simulations. The IEEE 30-bus test system is a simple approximate representation of the American power grid [39]. In this test system, there are 30 substations, 44 transmission lines and 6 generators. The IEEE 39-bus test system is the simplification of the well-known New-England area power grid [39]. This system includes 39 bus stations, 47 transmission lines and 10 generators. The IEEE 118-bus test system represents part of the US Midwestern power grid [39]. This system contains 118 substations, 54 generators and 186 transmission lines. In the preliminary test, the data of IEEE 39 and 118-bus test system obtained from MATPOWER are found to be inconsistent with the *N-1* criterion, so the capacity limits of several transmission lines are enhanced to comply with the *N-1* criterion in these systems. Table 2-1 shows the number of components in each system, where N_B is the number of transmission lines, N_L is the number of substations, N_D is the number of load buses, and N_G is the number of generators.

Table 2-1 The number of components in the adopted test system

Test system	$\boldsymbol{N_B}$	$\boldsymbol{N_L}$	$\boldsymbol{N_D}$	$\boldsymbol{N_G}$
IEEE 30- bus system	30	44	20	6
IEEE 39-bus system	39	47	21	10
IEEE 118- bus system	118	186	99	54

The load rate of a power system has a measurable impact on the process of cascading failures. Here, the load rate is taken into consideration in the simulation, and the load demand changes can be calculated as:

$$D_L(\beta) = \beta * D_{LS} \tag{2.6}$$

where D_L is the actual load demand level, D_{LS} is the standard load demand level obtained from [39], and β represents the load level of each simulation.

The performance of the proposed multi-component attack strategies is discussed in the following. The attacked components can be identified by the proposed two critical component identification methods. Transmission lines are chosen as the target component based on the following considerations. A modern power grid is composed of a large number of transmission lines, generators and substations. Typically, large-capacity generators and critical substations are relatively more heavily protected according to the safety and security measures. For transmission lines, their spanning length ranges from dozens of miles to hundreds of miles, and most part of the transmission lines is exposed to the wild environment, which makes it unrealistic to protect each portion of the transmission line. However, if one portion of a transmission line is disconnected, the line will lose its function immediately. Therefore, transmission lines are easy targets for attackers especially for physical attackers. The proposed multi-component attack strategies also apply to other power system components such as substations and generators.

The simulations are conducted based on the cascading failure simulator (CFS). At the beginning of the simulation, to identify the initial tripped transmission lines, static strategy and dynamic strategy are used; meanwhile, the edge betweenness centrality and electrical node significance are adopted as the indices of line criticality.

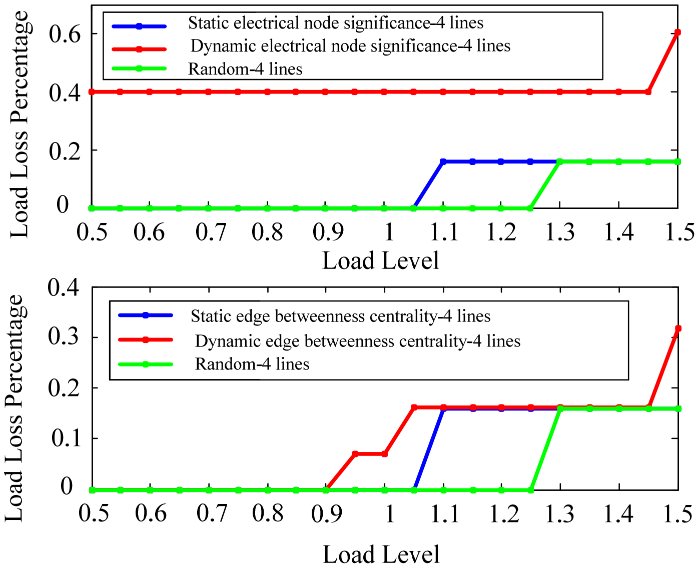


Figure 2-3 Comparison of static strategy and dynamic strategy in IEEE 30-bus system

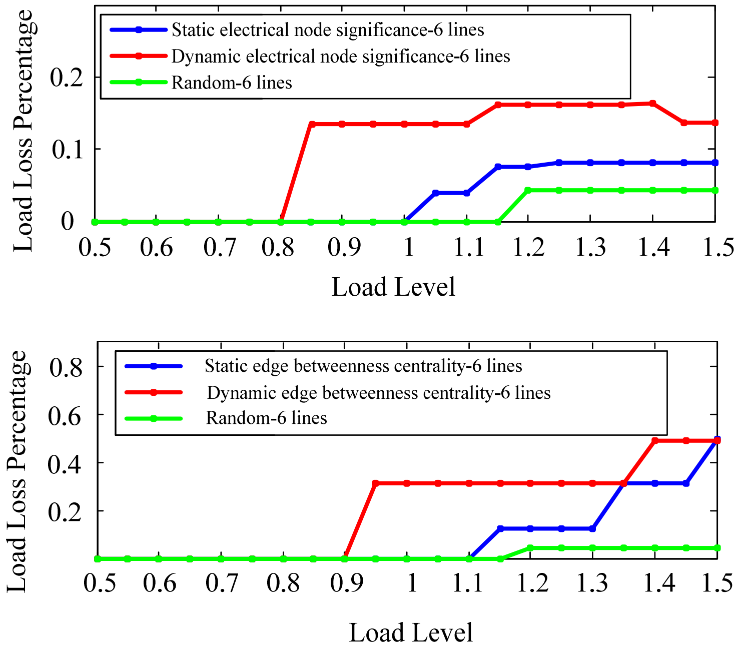


Figure 2-4 Comparison of static strategy and dynamic strategy in IEEE 118-bus system

Figure 2-3 and Figure 2-4 show the comparison of the results of static strategy and dynamic strategy, respectively; and the effectiveness of random attack is also shown here as a reference. Here, “Random-6” represents attacking six stochastically generated lines.

As can be seen from the figures, when the number of attacked lines and the line criticality index are the same, the consequences caused by the dynamic strategy is more serious than the static strategy, while both of them have led to more severe consequences than the random attack strategy.

2.4 Conclusions and Future Work

In this chapter, the cascading failures initiated by multi-component attacks are studied from the perspective of power system security. Static strategy and dynamic strategy are examined, and two criticality indices including the edge betweenness centrality and electrical node significance are adopted. Simulations are carried out based on the IEEE 30-bus test system, IEEE 39-bus test system and IEEE 118-bus test system, and the effectiveness of the proposed strategies is studied and compared.

This work may provide a new horizon for the prevention and mitigation of cascading failures caused by multi-component attacks. In practice, if the critical components are properly identified, then corresponding strategies could be developed and deployed to prevent or mitigate a cascading failure and ultimately prevent a probable large-scale blackout of electric power systems.

Chapter 3 Space-Pruning Enumerative Search Strategy for Identifying Critical Combinations of Multiple Same Types and Different Types of Components

3.1 Introduction

In chapter 2, static strategy and dynamic strategy are proposed to identify critical components in a power grid. The number of identified critical combinations is limited, which limits its ability to fully reflect the vulnerability level of the power grid. So, more effective methods should be developed to identify a more comprehensive set of critical component combinations.

Electric power systems are comprised of a large number of different components, including transmission lines, substations, transformers, generators and so on. In chapter 2， combinations of multiple same types of components are considered. Actually, the combination of different types of components can be targeted too. Some preliminary research has been conducted in this field. In [14], the joint line-substation attack scenario is investigated. In this chapter, the probable vulnerability by attacking multiple different types of components is studied.

The most direct method to find out the most critical component combination is the exhaustive enumeration; however, usually it is computationally unacceptable for a large power system. Here let’s take the IEEE 118-bus network as an example. On average, it takes about 0.1 seconds to run a cascading failure simulation using the cascading failure simulator on a computer with an i5-3230M processor, then it would take about fifty-six days to enumerate all the four-line attack results on this computing platform.

If the search space can be reduced significantly, the enumerative search strategy can be implemented to identify the critical component combinations. Specifically, the criticality of the component can be found based on the enumeration method if the number of components to be searched is reduced. The calculation time of such an enumeration method would be feasible especially for a power system which is not very large.

Furthermore, a space-pruning enumerative search strategy for identifying the critical combinations of multiple same types and different types of components will be proposed. The main idea of the space-pruning enumerative search strategy is to first get the criticality of each potential target component from low-order enumeration data, then the components will be sorted according to their criticality and the components with the greatest criticality will be selected to form a new search space; finally the enumerative search will be conducted in the new search space to find out critical component combinations.

The remainder of this chapter is organized in the following way. The proposed search space pruning enumeration strategy for identifying critical combinations of multiple same types and different types of components are presented in the subsequent sections of 3.2 and 3.3 respectively. Case studies and simulation results are given in section 3.4. The conclusions and future work of this chapter are presented in section 3.5.

3.2 Space-Pruning Enumerative Search Strategy for Identifying Critical Combinations of Multiple Components of Same Types

In this section, the space-pruning enumerative search strategy for identifying the critical combinations of multiple same type of components is proposed. In general, seven steps are required to identify the critical components. The flow chart of the strategy is shown in Figure 3-1.

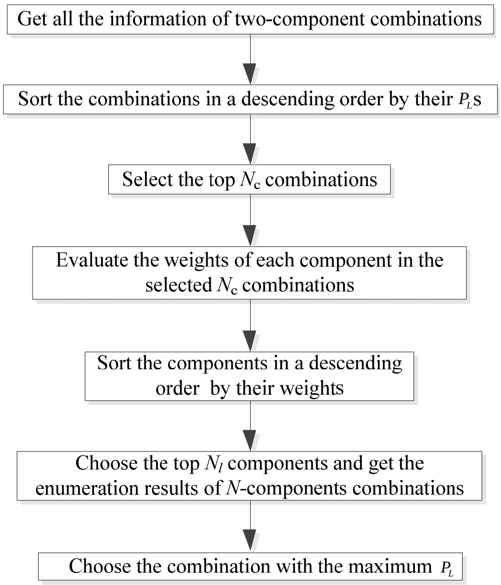

Get all the information of two-component combinations
Sort the combinations in a descending order by their P_Ls
Select the top N_c combinations
Evaluate the weights of each component in the selected N_c combinations
Sort the components in a descending order by their weights
Choose the top N_l components and get the enumeration results of N-components combinations
Choose the combination with the maximum P_L

Figure 3-1 The flow chart of space-pruning enumerative search strategy

The two-component enumeration attack data of the target power network needs to be collected as the first step of the algorithm. Although it takes some time to get the two-component enumeration data, the computing time is acceptable even when the scale of the target system is large. Then the obtained combinations will be sorted in a descending order by their P_L, and the top N_C combinations will be chosen and used for weight calculation for the components. The value of N_C is decided by attackers and depends on the size of the target grid. In general, the value of N_C increases with the grid size.

The components from the selected N_C combinations will be analyzed and the weight will be calculated. The weighted algorithm is illustrated as follows: all the selected N_C combinations will be examined and any combination containing that component will be picked out. The weight of the component represents the number of these combinations. After the weight of each component is obtained, all components will be sorted in a descending order by their weights.

Among the sorted components, the top N_l components are selected as the components to be enumerated, and the *N*-components enumerative search will be conducted among them. After getting the enumerated results, the combination with the highest P_L will be selected which is considered as the most effective attack combination, and the components forming this combination are regarded as critical components. The newly defined search space comprised of N_l selected components is termed the pruned search space. In the pruned search space, the enumerative computation burden is significantly reduced.

3.3 Space-Pruning Enumerative Search Strategy for Identifying Critical Combinations of Multiple Different Types of Components

In this portion of the chapter, we investigated the strategy for attacking multiple different types of components based on the search space pruning enumeration technique. Transmission lines and generators are important parts of the power system. Here, the combinatorial line-generator attack strategy (*LGCAS*) is investigated as an example of the attack strategy targeting multiple different types of components.

A line-generator combination is comprised of a collection of lines and generators. If a line-generator combination, which consists of m transmission lines and n generators, is chosen as the target, such an attack model can be denoted as $LGCAS^{m-n}$.

When a line-generator combination is attacked, the transmission lines and generators in the collection will be disconnected from the power grid. The attack can be initiated by physical attacks, cyber-attacks or a combination of both. The transmission lines and generators in the effective attack scheme is considered to be critical transmission lines and generators.

For the attackers， the simplest way to get the effective attack schemes is to conduct a brute-force enumeration. For simplification, the combinatorial line-generator attack strategy based on brute-force enumeration is denoted by $LGCAS_E$, the subscript letter indicates the brute-force enumeration strategy. Even though the best attack schemes can always be found based on the enumeration method in theory, usually it is not viable because the enumeration strategy is resource demanding and prohibitively costly if the target system is large and the number of possible combinations is massive.

This problem is very prominent even in a power system which is not big. Take IEEE 118-bus system as an example, on average it takes about 0.1 seconds to run a cascading failure simulation on a computer with an i5-3230M CPU, so it would take about sixty-six days to collect all the $LGCAS_E^{3-1}$ simulation results to identify the effective attack schemes. Hence, enumeration strategy is not practical for this problem, not to mention the fact that the actual power grid may be far greater than the IEEE 118-bus system.

3.3.1 Joint Line-Generator Attack Approach based on Reduced Search Space

A joint line-generator attack strategy based on the pruned search space can be proposed in this section. Such a strategy is termed $LGCAS_{RE}$, the subscript “RE” means reduced search space. When adopting $LGCAS_{RE}$ to identify effective attack combinations, several steps are required. The procedures of $LGCAS_{RE}$ are shown in Figure 3-2.

For the attackers, they usually have some expectation on the attack consequence before launching an attack. For instance, if the attackers want to make the whole power grid lose power, the expected value of P_L is one.

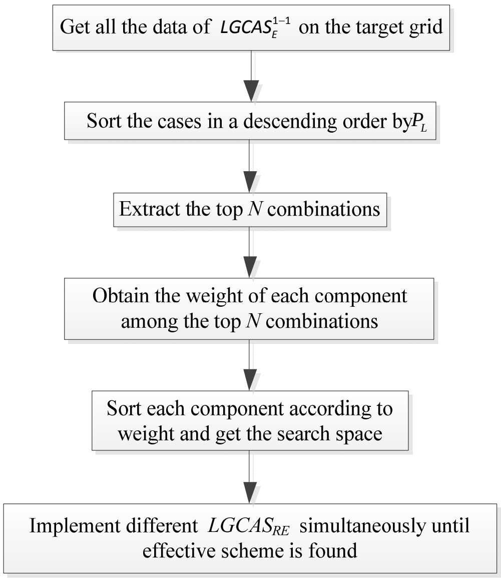

Get all the data of $LGCAS_E^{1-1}$ on the target grid
Sort the cases in a descending order byP_L
Extract the top N combinations
Obtain the weight of each component among the top N combinations
Sort each component according to weight and get the search space
Implement different $LGCAS_{RE}$ simultaneously until effective scheme is found

Figure 3-2 Procedure of $LGCAS_{RE}$

At the beginning of the procedure, the results of $LGCAS_E^{1-1}$ need to be collected. The computational burden of this step is tolerable. Next, the combinations of $LGCAS_E^{1-1}$ are sorted in a descending order by P_L, and the best N combinations are extracted. The value of N is set by attackers based on the scale of the power grid. In general, the value of N grows as the target grid size increases.

Among the N combinations, the transmission lines and generators will be identified and weighted. The weight of a transmission line is calculated as follows: For each identified line, all the combinations which contain the line will be examined in the N schemes, and the weight of this line is the sum of P_L in the identified scenarios. The meaning of the weight contains the number of occurrences in the chosen top N schemes along with its impact. Then the lines are sorted in a descending order by their weights. Calculating the weight of the generators follows the same procedure.

After the weights are obtained and the components are sorted, the top N_l sorted transmission lines and top N_g sorted generators are chosen independently. Since a generator with a larger capacity has higher potential to result in an effective attack, the credibility of critical generators is higher than the critical transmission lines. Thus the size of generator collections can be somewhat reduced to save the computing time.

The values of N_l and N_g need to be increased with the expansion of the grid size. While the generators in the power plants are generally well protected, transmission lines expand as wide as hundreds of miles with most of them being exposed to the wild, which makes transmission lines more vulnerable to attacks. Considering the attack difficulty, a $LGCAS_{RE}$ scenario, which consists of more lines or equal number of generators, will be considered. Different $LGCAS_{RE}$ simulations can be carried out simultaneously. For instance, to identify at least one attack scheme which can lead to the value of P_L above 0.7, $LGCAS_{RE}^{2-1}$, $LGCAS_{RE}^{3-1}$, $LGCAS_{RE}^{3-2}$, $LGCAS_{RE}^{4-1}$, etc. can be implemented simultaneously with a search space of reasonable size on a number of computers.

For a certain $LGCAS_{RE}$, search is conducted only among the selected N_l lines and N_g generators to identify the effective combinations. The search space formed by the N_l transmission lines and N_g generators is known as the reduced search space. In the reduced search space, the $LGCAS_{RE}$ computational time can be reduced significantly while a comprehensive set of the critical combinations can be obtained.

The $LGCAS_{RE}$ search can be further accelerated by dividing the sorted N_l lines and N_g generators into two symmetrical parts. The first part is composed of the lines and generators with higher weights, and the second part is composed of the lines and generators with lower weights. The higher weight lines collection is indicated by WL_H and the lower weight lines collection is indicated by WL_L. Similarly, WG_H represents the group of higher weight generators and WG_L represents the group of lower weight generators. The search space which is comprised of WL_H and WG_H is termed higher weight search space.

The higher weight search space will be explored first in the reduced search space during the process of $LGCAS_{RE}$ search. All the results will be recorded and processed. If one combination is found to satisfy the expectation, the simulation will be terminated and the qualified combination will be displayed.

If satisfactory results cannot be found in the higher weight search space, $LGCAS_{RE}$ search will be performed in the whole reduced search space except for the higher weight search space which has been searched. If critical combinations still cannot be found in the reduced search space, the search process will end and this $LGCAS_{RE}$ is seen as a void scenario.

It is worth noting that some critical combinations may be omitted based on this method; however, the result is still meaningful especially for attackers. In the actual case, since it is more difficult to attack a generator than a transmission line, an attack scheme composed of more lines and fewer generators could be a more effective option.

3.4 Simulations Results

The performance of the space-pruning enumerative search strategy is studied here. The space-pruning enumerative search strategy for both the same and different types of component attacks are simulated on IEEE 30, 39 and 118-bus systems.

3.4.1 A Case Study on Identifying Critical Combinations of Multiple Same Types of Components based on Enumerative Search

To demonstrate the space-pruning enumerative search strategy for identifying critical combinations of multiple components of the same type, in this section the enumerative search method is performed on the IEEE 30-bus benchmark. The case of multi-line attack is studied here, and all the results of three-line and four-line attacks are presented and discussed. For conducting further analysis, the collected data need to be appropriately processed. The procedure of data preprocessing is illustrated as follows.

At first, all the combinations of three-line and four-line will be sorted in a descending order by their P_L separately. Then, the top 30 schemes are extracted from the two kinds of combinations; the lines in the top 30 schemes will be picked out from the two kinds of combinations separately and the occurrence frequency (O_F) of each line will be calculated. Finally, the lines will be sorted in a descending order by their occurrence frequencies. The top twenty lines are shown in Table 3-1.

It can be observed that some lines occur frequently in both the top 30 three-line and four-line combinations. For example, the occurrence frequency of line 6 is 8 and 13 in the top 30 three-line and four-line combinations individually. It can be seen that the line criticality is not greatly impacted by the number of the target lines.

Table 3-1 The statistical data of the top 30 three/four-line combinations

Line number	three-line O_F	Line number	three-line O_F	Line number	four-line O_F	Line number	four-line O_F
21	14	20	3	7	21	15	3
16	10	19	3	20	16	11	3
7	10	18	2	39	15	44	2
6	8	17	2	6	13	24	2
34	7	8	2	19	9	23	2
24	7	5	2	21	7	22	2
15	5	39	1	8	6	10	2
25	4	33	1	5	6	34	1
11	4	31	1	30	4	33	1
23	3	9	1	16	3	31	1

3.4.2 Effectiveness of the Space-Pruning Enumerative Search Strategy for Multiple Homogeneous Components Outages

During the section, the proposed dynamic strategy is adopted for comparison. To compare the effectiveness of the space-pruning enumerative search strategy and dynamic strategy, the case of multi-line attack is studied. Simulations are conducted based on the

IEEE 39-bus system with the fixed base load profile. The simulation results are shown in Figure 3-4. The pruned search space N_l for attacking different lines is listed in Table 3-2.

Table 3-2 Pruned search space for attacking different lines in IEEE 39-bus system

Total line number	**47**			
Attacked lines	3	4	5	6
Pruned search space	10	15	15	15

In order to verify the effectiveness of the above attack strategies, the best results chosen from the results of the enumeration strategy are also shown. The results in Figure 3-3 clearly show that the performance of the space-pruning enumerative search strategy is better than the dynamic strategy. More notable observation is that the performance of space-pruning enumerative search strategy is highly close to the optimal results obtained from the enumerative search. This observation proves the validity of the space-pruning enumerative search strategy.

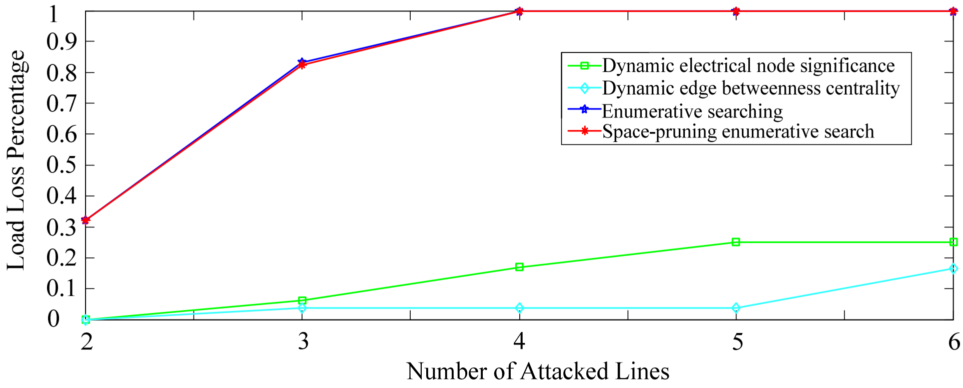


Figure 3-3 Comparison of the space-pruning enumerative search and dynamic strategies in the IEEE 39-bus system, the load level is fixed as 1.

Additionally, the space-pruning enumerative search strategy and the dynamic attack strategy adopt different criticality indices of lines for further testing. The test platform is the IEEE 118-bus system and the performances of two strategies are shown in Figure 3-4. The pruned search space N_l is listed in Table 3-3.

Table 3-3 Pruned search space for attacking different lines in IEEE 118-bus system

Total line number	186	
Attacked lines	3	6
Pruned search space	10	20

Considering it would take too long time to get the enumeration results, the enumerative search strategy is not implemented.

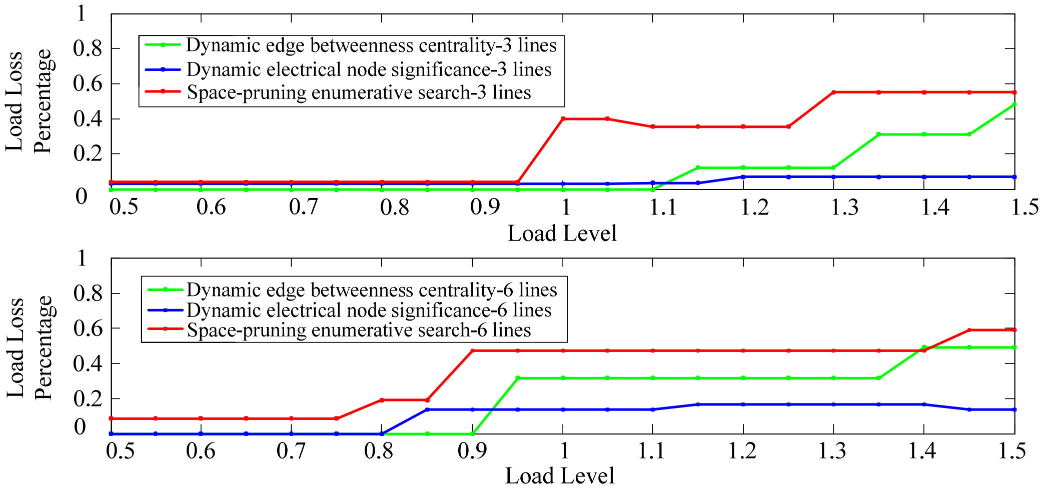


Figure 3-4 Comparison of the space-pruning enumerative search and dynamic strategies in IEEE 118-bus system, where the load level ranges from 0.5 to 1.5.

The results of Figure 3-4 prove that both edge betweenness centrality and electrical node significance are effective criticality indices when the power system operates at the rated state with a fixed base load profile. It can be seen that the performance of the space-pruning enumerative search strategy will surpass the dynamic attack strategy when the number of the target lines is large enough.

Also, it is clear that the power grid will become more vulnerable with the increase of load level, so that under the same attack mode, more serious damage could be caused. With the increase of power load level， more attention should be paid to the monitoring and protection function of the power system. For example, if the load level of a power system is high in a hot summer afternoon, power dispatchers should concentrate on the real-time state of the grid, and a contingency response plan for such a situation should be developed.

3.4.3 A Case Study of Multiple Different Types of Components Combinations based on Enumerative Search

To demonstrate the space-pruning enumerative search strategy for identifying the critical combinations of multiple different types of components, here the brute force search is conducted based on the IEEE 39-bus test system. The results of $LGCAS_E^{1-1}$, $LGCAS_E^{2-1}$, $LGCAS_E^{1-2}$ and $LGCAS_E^{3-1}$ are gathered and processed, and the flow of this process can be illustrated in Figure 3-5.

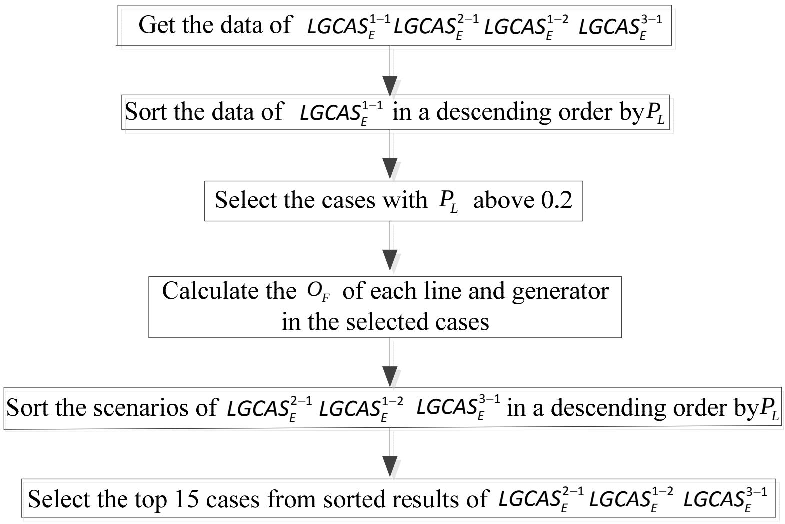

Get the data of $LGCAS_E^{1-1} LGCAS_E^{2-1} LGCAS_E^{1-2}$ $LGCAS_E^{3-1}$
Sort the data of $LGCAS_E^{1-1}$ in a descending order byP_L
Select the cases with P_L above 0.2
Calculate the O_F of each line and generator in the selected cases
Sort the scenarios of $LGCAS_E^{2-1}$ $LGCAS_E^{1-2}$ $LGCAS_E^{3-1}$ in a descending order byP_L
Select the top 15 cases from sorted results of $LGCAS_E^{2-1} LGCAS_E^{1-2}$ $LGCAS_E^{3-1}$

Figure 3-5 Method for processing the enumerated data

The main steps of the process procedure are illustrated as follows: At the beginning all the combinations in $LGCAS_E^{2-1}$, $LGCAS_E^{1-2}$and $LGCAS_E^{3-1}$ are sorted in a descending order by their P_L respectively; then for each $LGCAS_E$ the best 15 schemes are chosen. Table 3-4 shows the results.

Some observations can be made by analyzing this data. First, the transmission lines and generators occur in the high P_L value $LGCAS_E^{1-1}$ attack scenarios also have high occurrence frequencies in the effective $LGCAS_E^{2-1}$, $LGCAS_E^{1-2}$ and $LGCAS_E^{3-1}$ attack schemes.

Table 3-4 Case study of the enumerative search strategy

No.	$LGCAS_E^{1-1}$				$LGCAS_E^{2-1}$		$LGCAS_E^{1-2}$		$LGCAS_E^{3-1}$	
	line	O_F	***generator***	O_F	***schemes***	P_L	***schemes***	P_L	***schemes***	P_L
1	1	1	31	3	35, 39-33	0.8328	39-33, 34	0.8328	35, 38, 40-33	1
2	11	1	32	3	34, 39-34	0.8328	36-33, 34	0.8316	34, 38, 40-34	1
3	12	1	33	3	35, 36-33	0.8316	39-31, 32	0.8241	34, 36, 39-34	1
4	14	1	34	1	34, 36-33	0.8316	36-31, 32	0.8241	23, 39, 43-32	1
5	16	1	35	1	20, 39-31	0.8241	23-35, 36	0.6858	23, 36, 43-32	1
6	17	1	36	1	20, 36-31	0.8241	1-35, 39	0.6714	23, 36, 43-31	1
7	20	1	37	2	14, 39-32	0.8241	1-36, 39	0.6543	17, 23, 36-38	1
8	32	1	38	3	14, 36-32	0.8241	20-31, 39	0.6476	17, 23, 39-38	1
9	34	1	39	14	2, 39-39	0.7642	14-32, 39	0.6476	14, 20, 36-38	1
10	36	6			2, 36-39	0.7642	39-31, 33	0.6270	14, 20, 39-38	1
11	38	1			1, 39-39	0.7642	36-31, 33	0.6270	11, 12, 36-38	1
12	39	7			1, 36-39	0.7642	41-35, 38	0.6230	11, 12, 39-38	1
13	40	1			9, 10-39	0.7519	39-32, 33	0.6230	4, 14, 36-32	1
14	42	1			3, 39-33	0.7429	36-32, 33	0.6230	1, 13, 39-32	1
15	43	5			10, 12-35	0.7314	17-32, 35	0.6115	18, 36, 43-32	1

Furthermore, in general, if a line or a generator occurs frequently in the top $LGCAS_E^{1-1}$ scenarios, it is more likely to appear in the top $LGCAS_E^{1-2}$, $LGCAS_E^{2-1}$ and $LGCAS_E^{3-1}$ scenarios. For instance, among all the lines which appear in the top 31 $LGCAS_E^{1-1}$ schemes, the lines with the highest occurrence frequency is line 36 and line 39. It can also be seen that line36 and line39 appear repeatedly in the top $LGCAS_E^{2-1}$, $LGCAS_E^{1-2}$ and $LGCAS_E^{3-1}$ schemes. It can be inferred from the observation that if the lines and generators do result in effective $LGCAS^{1-1}$ schemes, they are likely to constitute effective attack schemes when they are incorporated in the following $LGCAS$ attack schemes. Equally noteworthy are the low occurrence frequency lines in the top 31 $LGCAS_E^{1-1}$ schemes. They also have low frequency of occurrence in the top $LGCAS_E^{2-1}$, $LGCAS_E^{1-2}$and $LGCAS_E^{3-1}$ schemes.

These observations also apply to the generators with more obvious observations. In general, the occurrence frequency in the effective attack schemes of a generator is proportional to its generation capacity.

3.4.4 Demonstration of the Attack Strategy for Combinations of Multiple Different Types of Components

In this section, the performance of the multiple different types of components attack strategy is tested in the IEEE 30 and 118-bus test systems. The joint lines-generator attack is adopted as an example for illustrating the attack strategy targeting multiple different types of components.

The proposed $LGCAS_{RE}$ is tested. The simulation results and observations are shown in detail. The performance of the node attack strategy is also presented for comparison. The node attack strategy adopted here is based on the electrical node significance which reflects the criticality of the nodes for cascading failures in a power system [34], [35]. When attacking N_s nodes using the electrical node significance strategy, the electrical node significance of each node in the power grid needs to be calculated, then the nodes will be sorted in a descending order according to the electrical node significance. The top N_s nodes are selected as the targets.

In this part, a transmission line, a generator or a node is called a component. Specifically, when conducting $LGCAS_{RE}$, a component represents a transmission line or a generator, while a component means a node when referring to the node attack strategy.

Figure 3-6 shows the performance comparison of the two attack strategies. The reduced search space adopted here is $N_l = 15$ and $N_g = 5$. The range of the attacked node number is 2 to 6 for the electrical node significance based node attack strategy. The corresponding results of the $LGCAS_{RE}$ are attained from $LGCAS_{RE}^{1-1}$, $LGCAS_{RE}^{2-1}$, $LGCAS_{RE}^{3-1}$, $LGCAS_{RE}^{3-2}$ and $LGCAS_{RE}^{4-2}$ respectively.

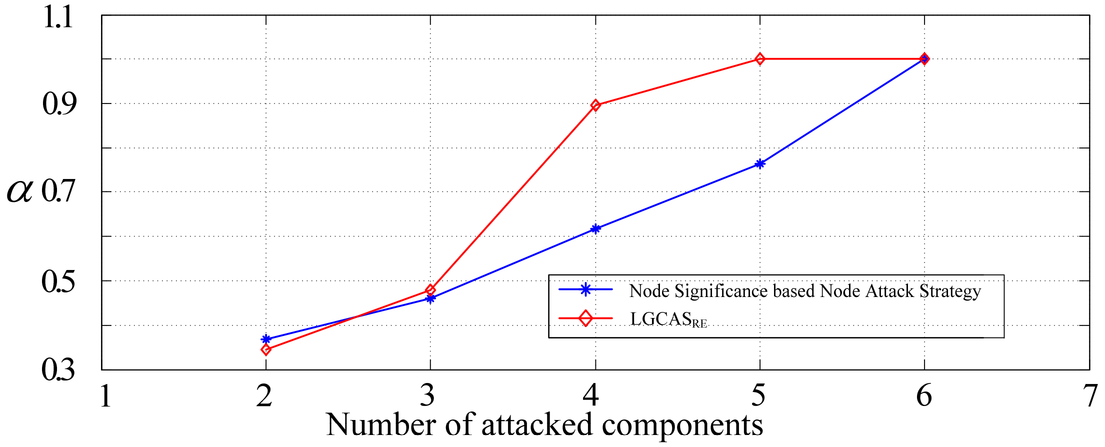


Figure 3-6 Comparison of $LGCAS_{RE}$ and the electrical node significance based node attack strategy and in IEEE 30-bus system

Figure 3-6 demonstrates that $LGCAS_{RE}$ outperforms the adopted node attack strategy in this case. For instance, $LGCAS_{RE}$ has salient advantages over the node attack strategy based on electrical node significance when the numbers of attacked components are same.

Since it is more difficult for the attackers to attack a node than to attack a single transmission line or a generator, $LGCAS$ is a more effective method for the attackers. The attackers need to carry out $LGCAS_{RE}$ to find effective $LGCAS$ attack schemes.

3.4.5 Effectiveness of the Search Space Reduction Algorithm for $LGCAS$ Attack Schemes

In this part, the capabilities of the $LGCAS_{RE}$ will be tested in the IEEE 118-bus test system. Initially, all the simulation results of $LGCAS_{E}^{1-2}$ are saved and processed for comparison. The required simulation time is 9.27 hours. Though the simulation time appears long, it is tolerable for such a complex problem. After obtaining the enumerated outcomes, all the combinations will be sorted in a descending order by the value of P_L. It can be seen from the enumeration data that there are 11 attack scenarios where P_L values are larger than or equal to 0.3090. These P_L values are depicted in Table 3-5.

Considering that among the enumeration results of $LGCAS_{RE}^{1-2}$ the best value of P_L is 0.429, the threshold here is set as 0.42 for $LGCAS_{RE}^{1-2}$. $LGCAS_{RE}^{1-2}$ is then implemented to identify an effective combinatorial attack strategy made up of one transmission line and two generators.

The top 100 sorted $LGCAS_{RE}^{1-1}$ combinations are chosen as the sampling space. The value differences of N_l, N_g are also considered. Table 3-6 shows the values of P_L which are above or equal to 0.309. In the following $LGCAS$, the attack scenarios whose P_L values are above or equal to 0.309 are defined as effective scenarios.

When $LGCAS_{RE}^{1-2}$ is implemented and the reduced search space is $N_l = 10$ and $N_g = 5$, three effective attack schemes can be found in the higher weight search space whose elapsed time is 1.53 seconds. When the search space is $N_l = 40$ and $N_g = 10$, six effective scenarios can be identified in the higher weight search space. The simulation time is about 18.68 seconds. This phenomenon shows that with the increase of the reduced search space, more effective attack scenarios can be identified.

Table 3-5 The processed data of $LGCAS_E^{1-2}$

Strategy	No.	P_L	Elapsed time
$LGCAS_E^{1-2}$	1	0.4279	9 hours and 16 minutes
	2	0.3900	
	3	0.3669	
	4	0.3669	
	5	0.3621	
	6	0.3423	
	7	0.3284	
	8	0.3250	
	9	0.3200	
	10	0.3116	
	11	0.3090	

Table 3-6 The processed data of $LGCAS_{RE}^{1-2}$ with different search spaces

Strategy	No.	P_L	Elapsed time
$LGCAS_{RE}^{1-2}$ $N_g = 10$ $N_g = 5$	1	0.4290	1.53 seconds
	2	0.3899	
	3	0.3200	
Strategy	No.	P_L	Elapsed time
$LGCAS_{RE}^{1-2}$ $N_l = 40$ $N_g = 10$	1	0.4290	18.48 seconds
	2	0.3899	
	3	0.3250	
	4	0.3200	
	5	0.3116	
	6	0.3090	

It should be noted that all the identified effective $LGCAS_{RE}^{1-2}$ results are found in the higher weight search space. The validity of higher weight search space can be verified through this observation.

Given that the enumerative search in the higher weight search space will stop if effective attack schemes can be identified, it is reasonable to believe that compared with the higher weight search space, more critical combinations can be identified in the reduced search space while the consumed time is not significantly increased.

3.5 Conclusions and Future Work

Concluding this chapter, a study of the attack strategy for multiple same types or different types of components based on the space-pruning enumerative search strategy has been carried out.

The IEEE 30-bus, 39-bus and 118-bus test systems are used for the simulation studies. In these test systems, the proposed attack strategy along with the reduced search space algorithm has turned out to be effective. The case of multi-line attacks is studied as an example of the space-pruning enumerative search strategy based attack strategy targeting the same types of components. The combinatorial line-generator attack strategy (*LGCAS*) is investigated as an example of the attack strategy targeting the multiple different types of components. To identify an effective joint line-generator attack strategy, the search space reduction based combinatorial line-generator attack strategy is proposed. This strategy can also be extended to identify other critical combinations of multiple different types of components.

In future studies, the space pruning method will be further improved. Prevention and mitigation methods can be developed based on this work to reduce the impact of multi-component attacks which may initiate disastrous cascading failures.

Chapter 4 Search Space Conversion and Reduction Strategy based Intelligent Search Method

4.1 Introduction

As previously stated, it is essential to identify critical component combinations on power system vulnerability analysis. Several methods can be adopted to identify critical component combinations. The simplest way is the brute-force enumeration method. Theoretically, all the critical combinations can be identified based on this method with the disadvantage of a heavy computational burden which is impractical especially for a large-scale power grid. As introduced in the previous two chapters, static strategy, dynamic strategy and space pruning enumerative search strategy produce the critical component combination identification method which can also be adopted to identify critical component combinations.

For the target problem, a power grid can be regarded as a black box in some sense. All possible target component combinations are the inputs and the critical component combinations can be seen as optimal solutions. The goal of critical components combination identification is to find out a few key combinations from a huge, high dimensional and nonlinear search space which makes it difficult for traditional identification methods to swiftly and comprehensively obtain critical combinations. The intelligent search algorithm is an effective approach to this problem. Reasonable solutions can be identified quickly in a large search space by using an intelligent search algorithm. Therefore, an intelligent algorithm is a good choice to identify the critical line combinations. Specially, an improved PSO algorithm is proposed for critical component combination identification in this chapter.

In order to further enhance the efficiency of intelligent search, a space conversion and reduction strategy based intelligent search method (SCRIS) is proposed. This strategy includes the following aspects: first the criticality of each component is obtained from the sampling results, then the components will be sorted according to their criticality and the components with the greatest criticality will be selected to form a new component space; finally intelligent search will be conducted in the new search space to identify critical component combinations.

The remainder of this chapter is organized in the following way. Section 4.2 introduces the effective component combination sampling method, and section 4.3 describes the weight calculation and new search space generation approach. Case studies and results are presented in section 4.4. The conclusions and future work of this chapter are given in section 4.5.

4.2 Effective Component Combination Sampling Method

A practical power grid is usually composed of thousands of substations and transmission lines. Let's assume there are R_N substations in a real-world power grid. For the binary encoding based intelligent search algorithms, in order to find critical node combinations, the size of the search space is 2^{R_N}, which would be an astronomical figure when R_N is large. Such a search space is too large for the traditional intelligent search algorithms.

The irregular distribution of critical component combinations in the search space makes the situation even worse. Limited by the performance of the traditional intelligent search algorithm, it is difficult for them to find the critical component combinations efficiently and effectively within an acceptable time scale.

Considering the number of the components in a power grid is assigned randomly, to reduce the difficulty of performing intelligent search, the number of components can be reassigned according to their criticality. In this way, a new search space will be generated in which the distribution of critical points is more concentrated as compared with the original search space. The performance of intelligent search algorithm could be improved by creating such a search space. It is worth pointing out that the criticality of the components does not need to be strictly precise, and the function of the converted space is to gather critical particles in a relatively concentrated area.

Although the sorted space is beneficial to improving the search performance, the search space is too large to identify critical component combinations for a large scale power grid. As mentioned before, in the resorted search space which is composed of the ranked components, there are reasons to think that the components with low weights play an insignificant role in the critical component combinations. Therefore, these components can be removed from the search space. As a result, in the search space composed of high weight components, the density of critical component combinations is much greater than that in the original search space.

As mentioned above, the validity of reordered search space is largely dependent on the criticality of the components. To obtain an “ideal” search space, the key issue is to find an effective way to sort the components so that the target combination can be restricted within a relatively small region that is close to the coordinate origin.

There are two major critical component sorting methods. One method is sorting the components according to the criticality index. In the case of line sorting, static strategy and dynamic strategy that are based on different criticality indices, like edge betweenness centrality, electrical node significance, and so on, can be adopted as sorting methods. The problem with this method is that the criticality of components may vary with the change of the simulation platforms.

Another way to get the criticality of each component is to analyze the sampling results. The characteristics of the simulation platform and the target power system have been taken into account during the sampling process. Several sampling methods are applicable to this problem, including random sampling, Monte Carlo sampling [40], Latin hypercube sampling [41] and Random Chemistry sampling [28], and so on. While there are multiple possible sampling methods to obtain the criticality of each component, the random chemistry method is the most powerful stochastic approach which is developed in [28] for quickly identifying the critical component combinations that will lead to cascading failures. Here, the Random Chemistry algorithm is adopted as the sampling tool.

Random chemistry algorithm

The random chemistry algorithm was originally proposed by Kauffman [42] in the applications of chemistry, and was further deployed by Eppstein and Hines [28] as a powerful stochastic tool for identifying critical component combinations in a power system. The main idea of this algorithm is to test the large randomly created component collection to find one combination that can initiate cascading failures, then to continuously reduce the size of the discovered critical combination by random search until a satisfactory combination is identified. The procedure of the Random Chemistry Algorithm can be described in the following steps [28].

Step1. Identify a large set of C_{ini} that can initiate cascading failures by testing randomly generated combinations from the target grid.

The set of potential target components in the target grid is denoted as set C_T, and the size of C_T is S_a. The set of initially selected target components in the power grid is denoted as set C_{ini}, and the size of C_{ini} is S. Before an attack is launched, attackers usually have an expected attack effect. Here a components combination which may cause consequences above a certain threshold value THR_{RC} is considered a critical combination which can initiate a cascading failure. At the beginning of the algorithm, S components will be randomly sampled from the S_a target components. The obtained set C_{ini} will be tested to see if it can initiate cascading failure. If set C_{ini} can cause cascading failures, it will be saved for step 2. If cascading failures cannot be initiated by set C_{ini}, the sampling process will be repeated for at most T_R times. If the desired set C_{ini} still cannot be found, the size of S will be doubled (the upper bound of S is S_a) and the above procedure will be repeated until a qualified set C_{ini} can be identified.

Step2. Reduce the selected set C_{ini} to set C_f with smaller size S_f by randomly sampling.

In this step, the size of the selected set C_{ini} will be further reduced to S_f based on the reduction factor R_F. R_F is a small real number greater than one. In the space reduction process, a subset C_f of set C_{ini} which contains $|S|/R_F$ (if $|S|/R_F$ is a decimal, $|S|/R_F$ is manipulated as an integer) components, will be tested to see if it could initiate a cascading failure. If a qualified subset C_f is identified, then the set C_{ini} will be replaced by C_f and the above process will be repeated until the size of set C_{ini} reaches S_f , where S_f is a small positive integer which limits the minimal size of set C_f. In any iteration, if attempts exceed T_R times, this RC trial will be considered to be invalid and another trial will start to identify the qualified combination.

Step3. Test possible subsets of C_f until the minimal critical combination is identified.

The size of set C_f obtained from step 2 is C_f. In this step, all the possible (S_f-1) component combinations from set S_f will be tested unless an (S_f-1) subset could cause cascading failure. This (S_f-1) subset will be used to find all the possible (S_f-2) combinations. This procedure will continue until an (S_f-2) combination produces a cascading failure. This step will be repeated until no smaller combinations could cause a cascading failure.

It should be noted that the performance of the RC algorithm is influenced by the specific type of power grids and the associated parameters [28].

When adopting a RC algorithm as the sampling tool, the sampling process can be introduced as follows: RC trials will be conducted for a reasonable number of times to obtain the critical component combinations. To find out a critical combination, the CFS simulation times vary in different RC trials. For purposes of conducting comparisons at the same scale, the sampling time is measured by the run times of the CFS, which is termed as T_{CFS}. If the attackers have the ability to attack at most N_A components, the RC algorithm S_f will be set as N_A. The obtained sampling results may include N_A component combinations, $N_A - 1$ component combinations, and until 2-component combinations.

4.3 Weight Calculation and New Search Space Generation Approach

Component Weight Calculation

The weight of each component will be calculated in the following way after the sample is obtained. Among the obtained samples, the combinations which can cause consequences above a certain threshold will be selected for weight calculation. The components occur in these combinations are weighted according to the following method:

For each of the components that appear in the selected combinations, all the combinations will be examined and any combination containing that component will be extracted. The weight of the component represents the number of combinations. After the weight of each component has been obtained, all components will be sorted in a descending order by their weights. Since it is possible that each component will not occur in the sampled combinations, the components which don't appear in the sampled combinations will be sorted according to their original index, and then appended to the ranked components. After the weight of each component has been obtained, all components will be sorted in a descending order by their weights. A new search space will be generated based on the sorted components.

Component Space Pruning Strategy

The search performance can be enhanced in the sorted search space; however, in a large scale power grid, the search space is still too huge and the performance of the intelligent search algorithm is limited by such a vast search space. As mentioned before, in the resorted search space which is composed of the ranked components, there are reasons to believe that the components with low weights play an insignificant role in the critical component combinations. Instead, in the search space composed of high weight components, the density of critical component combinations is much greater than that in the original search space. Assume that the number of the original components is T_O, the number of the selected components with top weights is T_T. Most critical component combinations will be included in the search space comprised of T_T components with moderate size. The component space pruning index S_R is defined here:

$$S_R = \frac{T_T}{T_O} \tag{4.1}$$

The reasonable value of S_R can be determined by the conclusion in [28], that is, the critical components make up only a small portion of the overall components. When considering five or less component combinations, desirable results can be achieved when the value of S_R is 10%. This value also takes the inaccuracies of component sorting into consideration. The search space will shrink dramatically based on the pruned components.

Improved PSO Algorithm

After the converted and reduced space has been obtained, intelligent search algorithms can be implemented in the new search space to identify critical component combinations, including the taboo search algorithm [43], particle swarm optimization (PSO) algorithm, genetic algorithm, simulated annealing algorithm, and so on. Here, an improved particle swarm optimization (PSO) algorithm is adopted due to its efficient search ability.

The Particle Swarm Optimization (PSO) algorithm is a well-known evolutionary algorithm which was proposed by James Kennedy and Russell C. Eberhart in 1995. The algorithm was inspired by the movement mechanism of a group of birds that are seeking food. The main idea of this algorithm is to generate a certain number of particles and to define an objective function first, then in each iteration, the fitness of the particles will be calculated and compared to generate the best local and global particle. The movements of these particles are guided by the position of the best local and global particles, respectively. The performance of this method is heavily dependent on associated parameters (e.g., iteration time, group size, inertia weight, etc.) and the design of the fitness function.

Derived from the basic particle swarm optimization algorithm, an improved PSO algorithm is developed for identifying critical component combinations more effectively. The following procedure describes the improved PSO algorithm:

Step1. Algorithm initialization

In this approach, the dimension of each particle is determined by the number of target components, each dimension corresponding to one target component. For example, if the attackers want to attack *J* transmission lines, then the particle's dimension is *J*. For each dimension, the searching range is determined by the number of potential target components. Assuming there are P potential target components in the power grid, the virtual distance of each dimension is [1, P] with interval 1. In the improved PSO algorithm, the search space has been greatly reduced. If the number of target components is *J*, the total search space is P^J. The speed of the particles are initiated randomly between $[V_{min}\ ,\ V_{max}]$, where V_{min} and V_{max} are the upper and lower bounds of the speed.

Step2. Fitness calculation and the best local and global position update

The position of each particle represents a target component combination. After initialization for each particle, the position will be sent to the cascading failure simulator as target components. The fitness function can be a different system failure definition, which is obtained from the simulation results of the CFS, like network-separation rate [28], net-ability drop percentage [44], load loss percentage (P_L), and so on. The goal of the particles is to find out the maximum fitness value. After obtaining the fitness value of each particle, the best local particle and the best global particle of this iteration will be determined. Here P_L is adopted as the fitness function.

Step3. Velocity and position update

The velocity and position of each particle are determined by the following formula [45]:

$$\begin{cases} v_{k-d}^{n+1} = wv_{k-d}^{n} + c_1 r_1^n (P_{b-d}^{n} - x_{k-d}^{n}) + c_2 r_2^n (P_{g-d}^{n} - x_{k-d}^{n}) \\ x_{k-d}^{n+1} = x_{k-d}^{n} + v_{k-d}^{n+1} \\ k = 1, 2, ..., M \\ d = 1, 2, ..., D \end{cases} \tag{4.2}$$

where w is the inertia weight, c_1 is the cognitive parameter and c_2 is the social parameter. M is the number of particles in the swarm and D is the dimension of a particle. r_1 and r_2 are random numbers between 0 and 1. v is the particle velocity and x is the particle position. P_b is the position of the best local particle and P_g is the position of the best global particle. The superscript n indicates the number of iterations. The subscript "*k-d*" represents the *d*-th dimension of the *k*-th particle, "*b-d*" represents the *d*-th dimension of the best local particle and "*g-d*" denotes the *d*-th dimension of the best global particle. During the iterations, if the particle's position exceeds the upper or the lower bound, the particle's position will be generated randomly. If the particle's velocity exceeds the upper or the lower bound, the speed will be maintained at the upper or lower bound.

Step4. Determining whether the algorithm should be halted.

The process of the algorithm will be terminated if one of the following termination conditions is satisfied:

1) The maximum number of iterations has been reached
2) No new critical combinations can be identified in 50 iterations

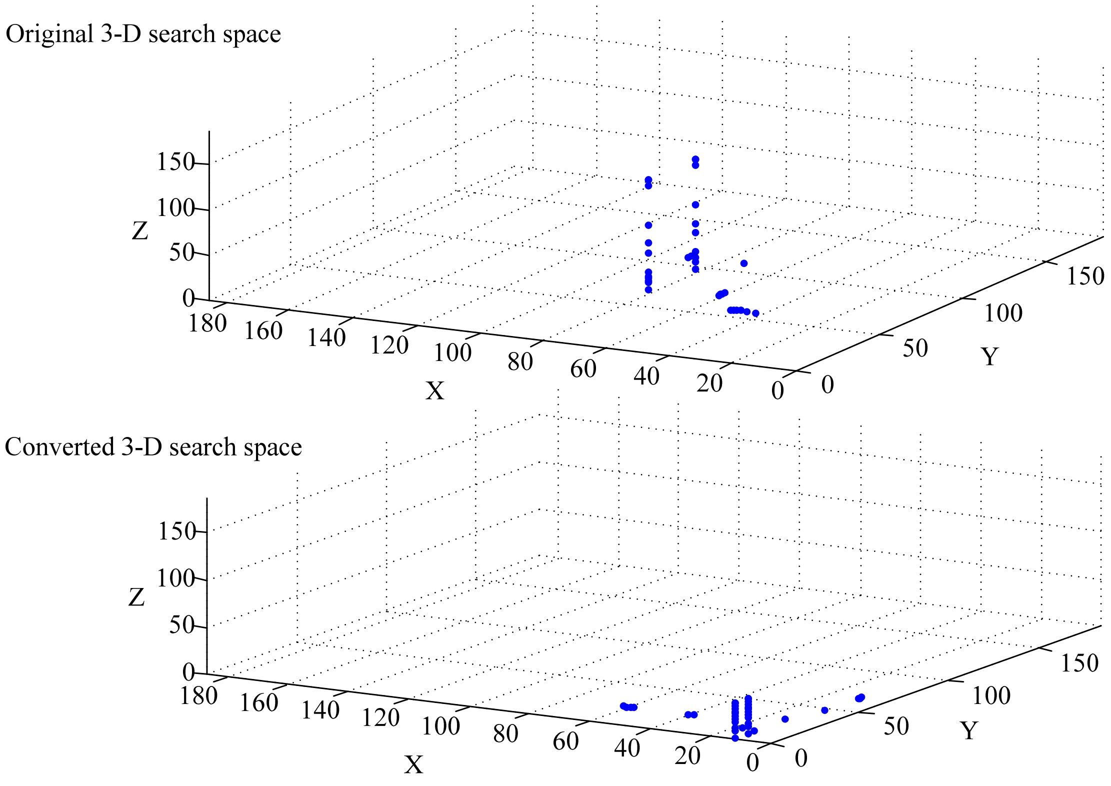
Before fulfilling any termination criterion, step 2 and step 3 will be repeated.

4.4 Simulation Results

In this section, the performance of the proposed space conversion and reduction strategy based intelligent search method (SCRIS) is test on IEEE 118 and 2383-bus systems based on MATLAB, the CFS is adopted as the cascading failure simulation tool. The main simulation results and conclusions are presented in the following subsections.

4.4.1 An Example of Space Conversion

Here is a simple example to demonstrate the effectiveness of the reordered search space.



Original 3-D search space

Converted 3-D search space

Figure 4-1 The comparison between original search space and converted search space

Figure 4-1 shows the comparison between original search space and converted search space. The top 40 three-line combinations extracted from the enumeration results are displayed in the two search spaces. It can be seen that the distribution of the top 40 particles in the resorted search space is obviously more concentrated, which means it is easier for the intelligent search algorithm to identify the critical particles. It can be expected that this advantage will be more evident for a large power grid considering multi-component combinations.

4.4.2 Search Space Comparison

In this part, the search spaces of the binary encoding method and the virtual distance encoding method are compared.

Search Space Comparison between the Binary Encoding Method and Virtual Distance Encoding method in the IEEE 118-bus system

The case of the line combinations attack is studied here. The search space of the binary encoding method and the 10-line combinationvirtual distance encoding method is listed in Table 4-1.

Table 4-1 Search space comparison between the binary encoding method and virtual distance encoding method on IEEE 118 bus system considering different numbers of target components

The number of target elements Encoding method	3	5	10
Binary encoding search space	$2^{186} = 9.8*10^{55}$		
Virtual distance encoding search space	$186^3 = 6.4*10^6$	$186^5 = 2.2*10^{11}$	$186^{10} = 4.9*10^{22}$

It can be seen from Table 4-1 that in the IEEE 118-bus system, the search space of the virtual distance encoding method is much smaller than that of the binary encoding method. For the binary encoding method, the search space is fixed for a specific system. In a power system, if the number of potential target components is T_C, the size of the search space is 2^{T_C}, which is an astronomical figure even for a not-so-large power system like the IEEE 118-bus system. As for the virtual distance encoding method, the search space grows as the number of target components increases; the growth rate is the number of potential target components. For example, when identifying effective 4-line combinations in the IEEE 118-bus system, the search space of the virtual distance encoding method is 118^4, while the search space is 118^5 when identifying effective 5-line combinations. In practice, a small portion of critical components can lead to serious consequences. Take IEEE 118-bus system as an example, serious damage could be caused if ten transmission lines are compromised. Considering this fact， the number of target components is usually a single digit, so that the search space of the virtual distance encoding method is limited to a very small range compared with the binary encoding method.

Comparison of Search Space Considering Attack Five Nodes in IEEE 30, 39, 118, 2383 bus systems

Table 4-2 shows the search space comparison between the binary encoding method and the virtual distance encoding method in different power systems. The case for searching effective five-line combinations is studied here. Practically, in IEEE 2383-bus system if five critical lines are compromised, over 90% of load shedding could be caused. Here, the space reduction ratio (SR_R) is defined as the ratio of the binary encoding search space and the virtual distance encoding search space.

It can be seen from Table 4-2 that even in the small IEEE 30-bus test system, the space reduction ratio is about 44, which means the virtual distance encoding search space is about 1/44 of the binary encoding search space. With the increase of power grid size, the space reduction ratio grows significantly. The virtual distance encoding search space is compressed drastically compared with the binary search space especially in a large-scale power grid, which is very favorable for conducting efficient intelligent search.

Table 4-2 Comparison of search spaces considering attack five nodes on IEEE 30, 39, 118, 2383-bus systems

Test System	**IEEE 30 bus test system**	**IEEE 57 bus test system**	**IEEE 118 bus test system**	**IEEE 2383 bus test system**
Binary Search Space	2^{30}	2^{57}	2^{118}	2^{2383}
Virtual distance encoding Search Space	30^5	57^5	118^5	2383^5
$\boldsymbol{SR_R}$	44.2	$2.4*10^8$	$1.5*10^{25}$	$2.9*10^{700}$

4.4.3 Performance of Space Conversion and Reduction Strategy

In this section, the performance of space conversion and reduction strategy is tested in the IEEE 118-bus system. The case of the line combinations attack is studied here. The random sampling method is adopted here as an example. 10,000 five-line combinations are stochastically sampled out of $C_{186}^{5} = 1.76 * 10^{9}$, the sampling rate is $\frac{10000}{C_{186}^{5}} = 5.7 * 10^{-6}$ which is obviously a very small rate. Among the sampled combinations, the combinations that can cause P_L above 0.5 are selected for line weight calculations. After obtaining the weight of each line, the top 50 lines are selected to form a new component space, $S_{T-R-10000-0.5}^{118-50}$. While the last 50 lines are selected to form another component space $S_{L-R-10000-0.5}^{118-50}$, the component space which is comprised of all the ranked 186 lines is $S_{F-R-10000-0.5}^{118-186}$. The improved PSO is implemented 50,000 times in $S_{T-R-10000-0.5}^{118-50}$, $S_{L-R-10000-0.5}^{118-50}$ and $S_{F-R-10000-0.5}^{118-186}$ respectively to search for effective 5-

line combinations. The parameters of the improved PSO is $\mathrm{w} = 1$, $c_1 = 1$, $c_2 = 1$, $V_{max} = 10$, $V_{min} = -10$, $\mathrm{nop} = 100$. Figure 4-2 shows the search results. The horizontal coordinate represents the attack effect. For example, 0.5-0.6 represents the attack effect's range: $0.5 \leq P_L < 0.6$; the vertical coordinate represents the number of combinations in each attack effect's range.

It can be seen clearly that in $S_{L-R-10000-0.5}^{118-50}$, no 5-line combination can be found to cause P_L above 0.1. Actually, among the 50,000 results obtained from $S_{L-R-10000-0.5}^{118-50}$, the highest P_L is 0.0165. The statistical data of search results obtained from $S_{L-R-10000-0.5}^{118-50}$ is shown in Table 4-3. This phenomenon indicates that in $S_{L-R-10000-0.5}^{118-50}$, almost no 5-line combinations has threatened the power grid. In other words, the transmission lines in $S_{L-R-10000-0.5}^{118-50}$ are not critical lines.

Table 4-3 Statistical data of search results obtained from $S_{L-R-10000-0.5}^{118-50}$

P_L	0-0.0165	0
Amount	4041	45959

The obvious comparison is the search results in $S_{T-R-10000-0.5}^{118-50}$. The number of combinations in each attack effect's range in $S_{T-R-10000-0.5}^{118-50}$ is far beyond the corresponding results in $S_{F-R-10000-0.5}^{118-186}$. It suggests that in $S_{T-R-10000-0.5}^{118-50}$, most critical combinations are included. It is worth pointing out that the space conversion and reduction strategy is effective even in the case of random sampling with a small number of samples. The performance of the proposed space conversion and reduction strategy can be improved if a more effective sampling method is adopted and more samples are collected.

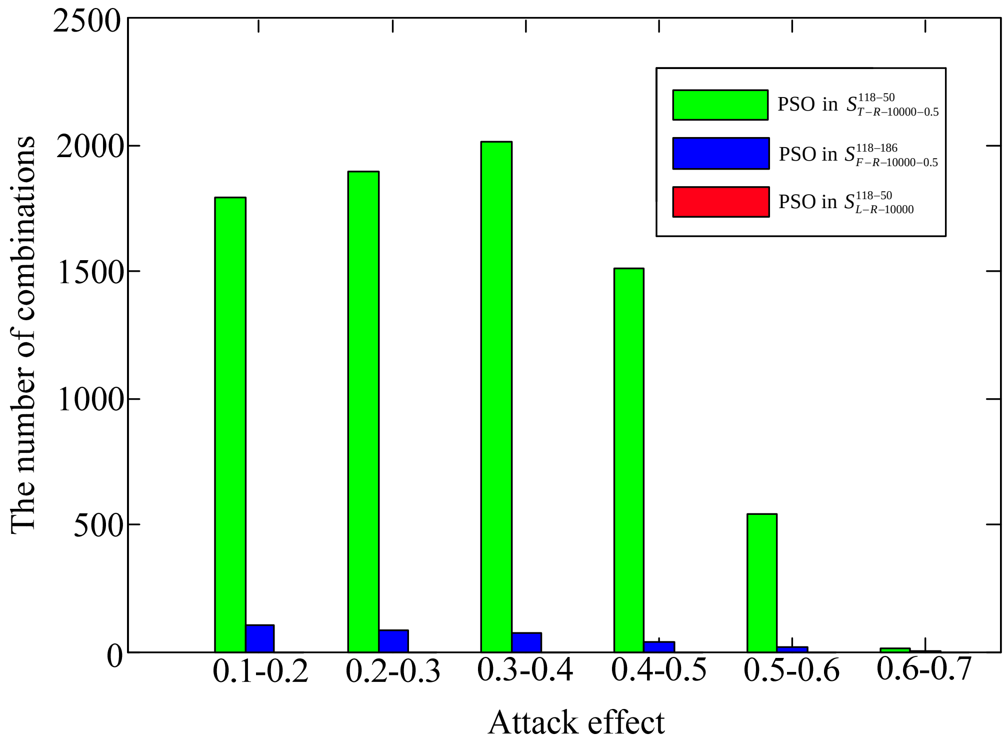


Figure 4-2 Search results of PSO in $S^{118-50}_{T-R-10000-0.5}$, $S^{118-50}_{L-R-10000-0.5}$ and $S^{118-186}_{F-R-10000-0.5}$

4.4.4 Performance Comparison between the Random Sampling Method and RC Sampling Method

In this section, the performance of the two sampling methods, that is, the random sampling method and RC sampling method, is tested on the IEEE 118-bus system and the 2383-bus system. The case of searching for effective five-line combinations is studied here.

Performance Comparison between the Random Sampling Method and RC Sampling Method in IEEE 118 bus system

The random sampling method and RC sampling method are tested based on the IEEE 118-bus system. For the random sampling method, 20,000 five-line combinations are stochastically sampled. Among the 20,000 sampled combinations, the combinations that can cause P_L above 0.5 are selected for line weight calculations. For the RC sampling method, THR_{RC} is set as 0.5.

Table 4-4 The top 50 sorted lines obtained from random sampling and RC sampling

Random sampling				**RC sampling**			
No.	Line	No.	Line	No.	Line	No.	Line
1	**8**	26	**39**	1	**7**	26	**34**
2	**36**	27	**102**	2	**9**	27	**37**
3	**33**	28	**42**	3	**8**	28	**80**
4	**38**	29	**70**	4	**36**	29	**160**
5	**104**	30	**83**	5	**31**	30	**55**
6	**9**	31	**32**	6	**33**	31	**140**
7	**21**	32	**29**	7	**19**	32	**50**
8	**7**	33	**11**	8	**21**	33	**174**
9	**96**	34	**3**	9	**96**	34	**61**
10	**107**	35	**73**	10	**38**	35	**74**
11	**31**	36	**152**	11	**32**	36	**172**
12	**97**	37	**120**	12	**17**	37	**143**
13	**51**	38	**61**	13	**51**	38	**89**
14	**93**	39	**90**	14	**54**	39	**15**
15	**37**	40	**62**	15	**22**	40	**138**
16	**17**	41	**48**	16	**141**	41	**18**
17	**5**	42	**88**	17	**16**	42	**121**
18	**18**	43	**67**	18	**76**	43	**117**
19	**127**	44	**50**	19	**48**	44	**120**
20	**141**	45	**82**	20	**183**	45	**90**
21	**131**	46	**69**	21	**110**	46	**62**
22	**126**	47	**26**	22	**109**	47	**108**
23	**154**	48	**179**	23	**178**	48	**27**
24	**142**	49	**54**	24	**142**	49	**30**
25	**95**	50	**89**	25	**105**	50	**99**

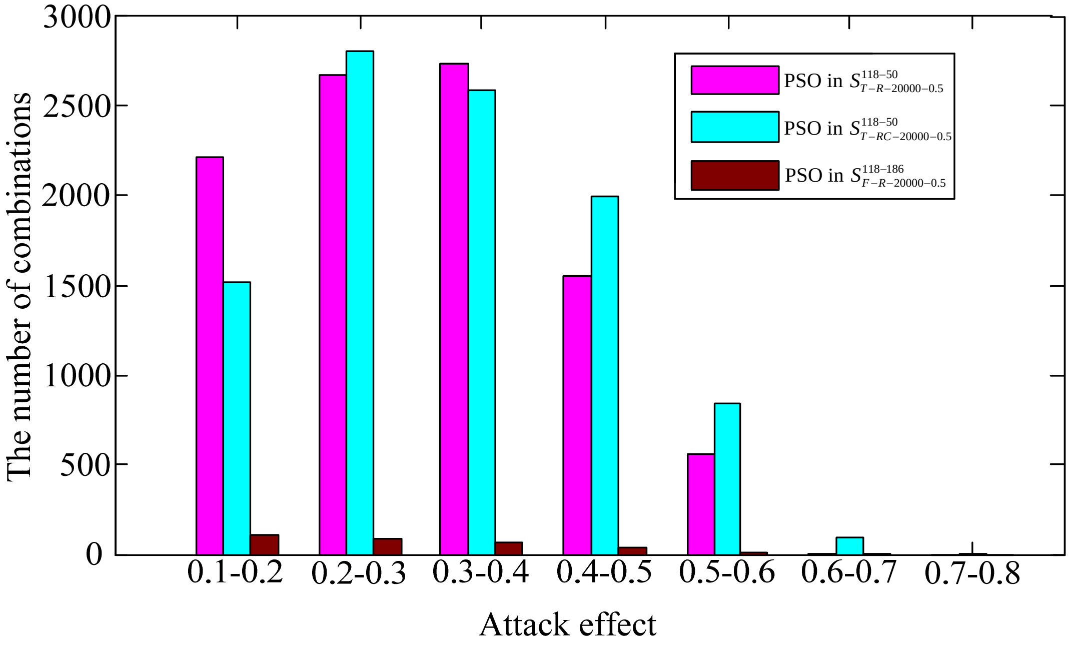


Figure 4-3 Search results of PSO in $S_{T-R-20000-0.5}^{118-50}$, $S_{T-RC-20000-0.5}^{118-50}$ and $S_{F-R-20000-0.5}^{118-186}$

After obtaining the weight of each line from the two sampling methods, the top 50 lines are selected respectively to form new component spaces, which are denoted as $S_{T-R-20000-0.5}^{118-50}$ and $S_{T-RC-20000-0.5}^{118-50}$. Table 4-4 shows the results of the top 50 sorted lines and the corresponding number obtained from the two sampling methods respectively. The component space consisting of all the ranked 186 lines obtained from the random sampling results is termed as $S_{F-R-20000-0.5}^{118-186}$.

PSO is implemented in $S_{T-R-20000-0.5}^{118-50}$, $S_{T-RC-20000-0.5}^{118-50}$and $S_{F-R-20000-0.5}^{118-186}$ respectively 50,000 times to acquire effective 5-line combinations. The parameters of the PSO are: $\mathrm{w} = 1$, $c_1 = 1$, $c_2 = 1$, $V_{max} = 10$, $V_{min} = -10$, $\mathrm{nop} = 100$. Figure 4-3 shows the search results.

It can be seen clearly from Figure 4-3 that in general, the number of critical combinations obtained from $S_{T-RC-20000-0.5}^{118-50}$ are greater than that of $S_{T-R-20000-0.5}^{118-50}$, especially when R_L is high, while the two above results are far better than the results obtained from $S_{F-R-20000-0.5}^{118-186}$. This results indicate that the RC sampling method is more effective than the random sampling method. As a result, in $S_{T-RC-20000-0.5}^{118-50}$, the distribution of critical lines is more concentrated so that critical combinations are more easily able to be searched out.

Performance Comparison of the Random Sampling Method and RC Sampling Method in IEEE 2383-bus system

In this section, the performance comparison between the random sampling method and RC sampling method is conducted in the IEEE 2383-bus system. The number of samples is 20,000. The sampling and weighting processes are the same as that discussed in the last section.

After obtaining the weight of each line, the top 50 lines are selected to form the component spaces, which are denoted as $S_{T-RC-20000-0.5}^{2383-50}$ and $S_{T-R-20000-0.5}^{2383-50}$.

The PSO is implemented in $S_{T-R-20000-0.5}^{2383-50}$ and $S_{T-RC-20000-0.5}^{2383-50}$ respectively 20,000 times to obtain the effective 5-line combinations. The parameters of PSO are： $\mathrm{w} = 1$, $c_1 = 1$, $c_2 = 1$, $V_{max} = 10$, $V_{min} = -10$, $\mathrm{nop} = 100$. Figure 4-4 shows the search results.

It can be seen clearly from Figure 4-4 that the search results in $S_{T-RC-20000-0.5}^{2383-50}$ have an evident advantage over the results in $S_{T-R-20000-0.5}^{2383-50}$. Compared with the results in the last section, in the medium-scale IEEE 118-bus system, the performance difference between the random sampling method and RC sampling method is not large. While in the larger-scale IEEE 2383-bus system, the performance of the random sampling method is poor in comparison with the RC sampling method. It can be concluded that, with the increase of the system size, the effectiveness of the random sampling method declines rapidly.

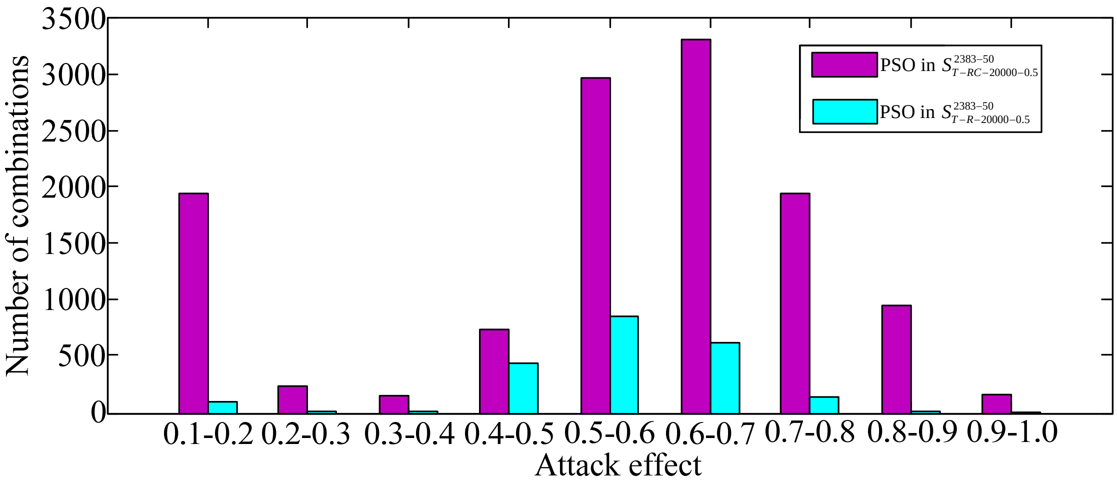


Figure 4-4 Search results of PSO in $S_{T-R-20000-0.5}^{2383-50}$ and $S_{T-RC-20000-0.5}^{2383-50}$

4.4.5 Performance Comparison between the Improved PSO based SCRIS and Random Chemistry

In this part, the performance comparison between the improved PSO based SCRIS and random chemistry is presented. The simulations are conducted in IEEE 118 and IEEE-2383 bus systems.

Performance Comparison between the Improved PSO based SCRIS and Random Chemistry Algorithm in IEEE 118 bus system

First the performance of the improved PSO based SCRIS and Random chemistry search algorithm is tested in IEEE-118 bus system. The case of searching effective five-line combinations is studied.

The adopted sampling method is the RC sampling method. The RC sampling parameters are $\mathrm{S} = 40$, $T_R = 20$, $THR_{RC} = 0.5$, and $T_{CFS} = 20{,}000$. The top 50 and 100 weighted lines are selected to form the component search space, $S_{T-RC-20000-0.5}^{118-50}$ and $S_{T-RC-20000-0.5}^{118-100}$. The improved PSO is implemented twice in the $S_{T-RC-20000-0.5}^{118-50}$ and $S_{T-RC-20000-0.5}^{118-100}$ 1,000,000 times respectively, searching for effective 5-line combinations. The parameters of the improved PSO are: $\mathrm{w} = 1$, $c_1 = 1$, $c_2 = 1$, $V_{max} = 20$, $V_{min} = -20$, $\mathrm{nop} = 50$, and $T_{CFS} = 1{,}020{,}000$. The random chemistry search algorithm is also executed to identify effective 5-line combinations. The parameters of the random chemistry search algorithm are: $\mathrm{S} = 40$, $T_R = 20$, $THR_{RC} = 0.5$, and $T_{CFS} = 1{,}020{,}000$. T_{CFS} of the PSO and RC are both 1,020,000 for the purpose of comparison under the same time scale.

Figure 4-5 shows the simulation results. As it can be seen clearly from the figure, the performance of the improved PSO based SCRIS has comprehensive advantages over the RC search. For further discussion, the number of identified combinations in each attack coincides with the range obtained from each method is listed in Table 4-5.

The search results obtained from PSO in $S_{T-RC-20000-0.5}^{118-50}$ have comprehensive advantages over the results attained from RC, and the performance of PSO in $S_{T-RC-20000-0.5}^{118-100}$ is similar to that of RC. For both RC and PSO in $S_{T-RC-20000-0.5}^{118-100}$, one five-line combination which can cause P_L above 0.8 is identified, while for PSO in $S_{T-RC-20000-0.5}^{118-50}$, no such a combination is identified. The above observations suggest that the search efficiency decreases with the elevation of S_R. Also, the five-line combination which can cause P_L above 0.8 is very likely to be excluded from $S_{T-RC-20000-0.5}^{118-50}$, which indicates that the efficiency and effectiveness of SCRIS are sensitive to the value of S_R. A reasonable S_R value can not only guarantee the search efficiency, but also ensures the quality of search results.

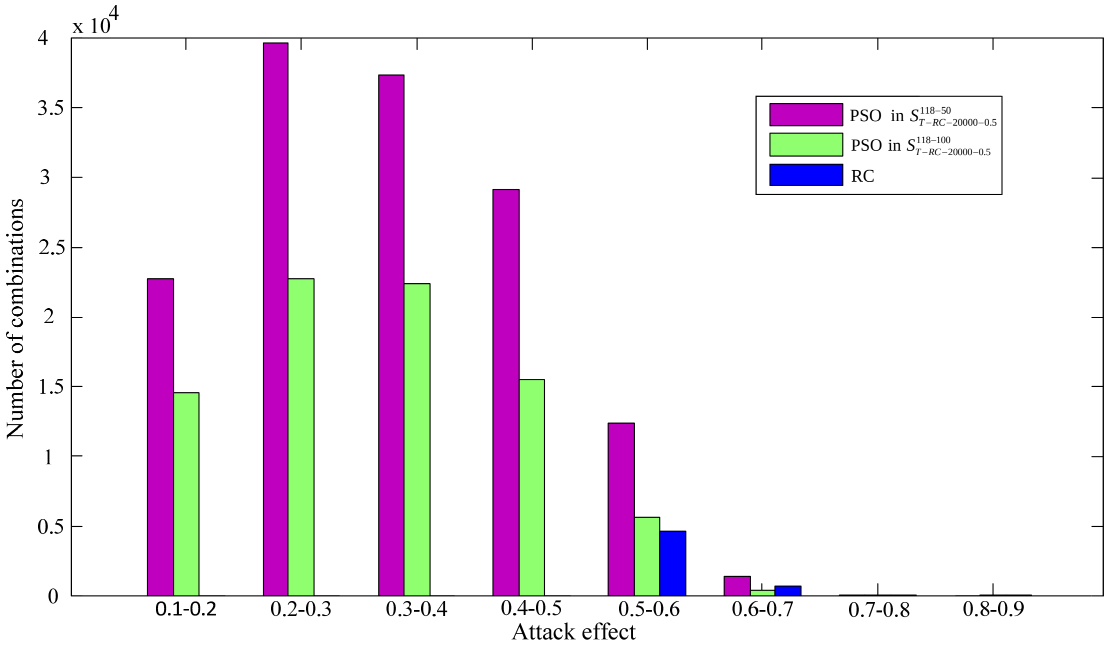


Figure 4-5 Performance comparison between the improved PSO based SCRIS and random chemistry algorithm in IEEE 118 bus system

Table 4-5 The number of identified combinations in each attack effect's range obtained from the improved PSO based SCRIS and random chemistry algorithm in IEEE 118 bus system

Search method Range of P_L	RC	PSO in $S_{T-RC-20000-0.5}^{118-100}$	PSO in $S_{T-RC-20000-0.5}^{118-50}$

0.5-0.6	4557	5644	12318
0.6-0.7	632	320	1378
0.7-0.8	16	13	72
0.8-0.9	1	1	0
0.9-1.0	0	0	0

The Improved PSO based SCRIS and Random Chemistry Algorithm in the IEEE 2383-bus system

In this section, the performance of the improved PSO based SCRIS and Random Chemistry search algorithm is tested in the IEEE-2383 bus system.

The sampling method for SCRIS is the RC sampling. The parameters are $\mathrm{S} = 80$, $T_R = 20$, $THR_{RC} = 0.5$, $T_{CFS} = 50{,}000$. The top 50 and 100 weighted lines are selected to form component space $S^{2383-50}_{T-RC-50000-0.5}$ and $S^{2383-100}_{T-RC-50000-0.5}$. Considering the cascading failure simulation on the IEEE 2383-bus system is time-consuming, the improved PSO is run for $S^{2383-50}_{T-RC-50000-0.5}$ and $S^{2383-100}_{T-RC-50000-0.5}$ 200,000 times. The parameters of the improved PSO are $\mathrm{w} = 1$, $c_1 = 1$, $c_2 = 1$, $V_{max} = 20$, $V_{min} = -20$, $\mathrm{nop} = 100$， and $T_{CFS} = 250{,}000$. The random chemistry search algorithm is also executed twice to identify the critical 5-line combinations. The parameters of the random chemistry search algorithm are $\mathrm{S} = 80$, $T_R = 20$, $THR_{RC} = 0.5$, and $T_{CFS} = 550{,}000$.

The simulation results are shown in Figure 4-6, and the number of identified combinations for each attack effect’s range obtained from each method is listed in Table 4-5. It can be seen from the simulation results that there are comprehensive performance advantages of the improved PSO based SCRIS with 250,000 times CFS simulations over that of RC search with 550,000 CFS simulations. The search efficiency of PSO in

$S^{2383-100}_{T-RC-50000-0.5}$ decreased to some extent compared with $S^{2383-50}_{T-RC-50000-0.5}$ due to the increase in the search space.

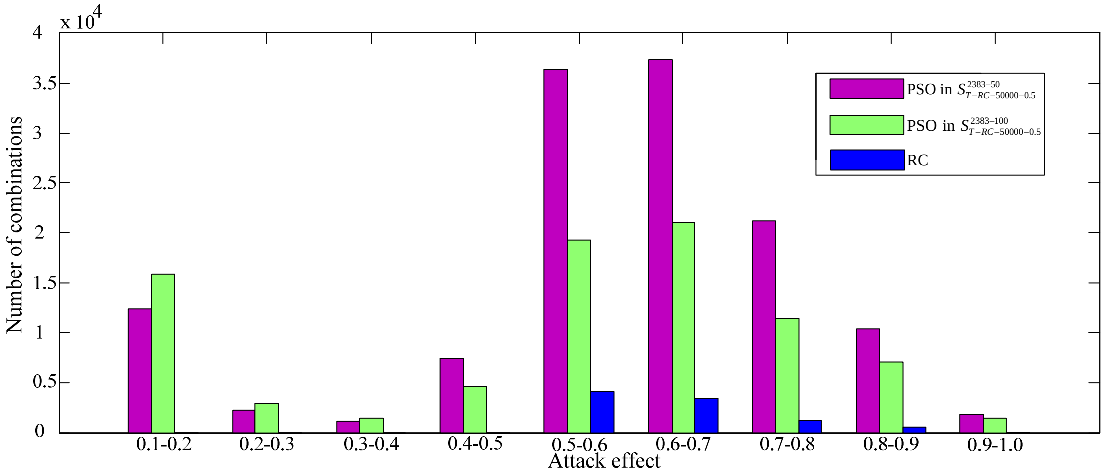


Figure 4-6 Performance comparison between the improved PSO based SCRIS and random chemistry algorithm in IEEE 2383-bus system

Table 4-6 The number of identified combinations for each attack effect's range obtained from the improved PSO based SCRIS and random chemistry algorithm in IEEE 2383-bus system

Search method Range of P_L	RC	PSO in $S^{2383-100}_{T-RC-50000-0.5}$	PSO in $S^{2383-50}_{T-RC-50000-0.5}$
0.5-0.6	4083	19282	36413
0.6-0.7	3451	21121	37361
0.7-0.8	1209	11388	21188
0.8-0.9	535	7029	10422
0.9-1.0	51	1441	1822

This observation indicates that the search efficiency and effectiveness of the improved PSO based SCRIS is much higher than what is obtained from the RC search algorithm in a large scale power grid. It can be concluded that the improved PSO based SCIRS is particularly suitable for quickly identifying critical component combinations in a large-scale power system.

4.5 Conclusions and Future Work

In this chapter, the space conversion and reduction strategy based intelligent search method (SCRIS) is proposed for efficiently and comprehensively identifying critical component combinations in a power grid. An improved PSO algorithm is adopted to further enhance the search ability.

The IEEE 118-bus system and the Polish 2383-bus system are used for simulations. The case of multiple line attacks is studied as an example of the SCRIS based attack strategy targeting the same types of components. In these systems, the proposed SCRIS as well as the improved PSO algorithm are validated to be effective. Based on this work, preventive methods may be developed to prevent multi-component attacks which could initiate cascading failures. This work also has a great significance in guiding the smart grid planning.

In future studies, the sampling method, weight calculation method and space reduction method will be improved to further enhance the performance of the SCRIS. Also, it is of importance to investigate the influence of S_R on the SCRIS search performance.

Chapter 5 Conclusion

In this thesis, the cyber-physical vulnerability of next-generation electric power systems in a smart grid environment is investigated for the purpose of performing power system cyber-physical security analysis considering the potential malicious attacks against power grids. The thesis proposes novel methods for effectively and comprehensively identifying critical component combinations which may initiate cascading failures. Based on the proposed methods, a holistic methodology could be established for enabling a comprehensive power system vulnerability analysis, which could provide practical solutions for the protection, operation, and planning of large-scale smart grids.

Based on the existing criticality indices, static strategy and dynamic strategy are here used to reflect the vulnerability of power systems. For different application purposes, more criticality indices could be developed and deployed in vulnerability evaluation.

Considering the diversity of power system components, this thesis also covers a study on attack strategies targeting multiple same or different types of components based on the space-pruning enumerative search strategy. In particular, system vulnerability due to multiple different types of components of power grid is carefully studied which however cannot be discovered by the conventional vulnerability analysis method developed for assessing the same types of components.

The space conversion and reduction strategy based intelligent search method (SCIRS) are also proposed for comprehensively and effectively identifying critical component combinations. Through this approach, the potential components are chosen and manipulated by the sampling strategy and space conversion and reduction strategy so that the search space shrinks to a smaller range. This technique significantly decreases the computational burden while having the ability to identify comprehensively vulnerable portions in the power grid. This method has advantages over traditional intelligent search approaches for cascading failure analysis. The improved PSO based SCIRS is proven to be particularly suitable for quickly identifying critical component combinations in a large-scale power system. Another significant advantage of SCRIS is that the algorithm is applicable to the power systems with different topologies as well as the diverse cascading failure simulation platforms.

Based on the proposed methods, preventive actions and response plans can be developed to diminish the effect of multi-component attacks which could initiate cascading failures. The results of search space conversion and reduction strategy based intelligent search method also have a great significance for decision-making support in achieving effective smart grid management.

References

[1] Liscouski, B., & Elliot, W. (2004). Final report on the august 14, 2003 blackout in the united states and canada: Causes and recommendations. *A report to US Department of Energy*, *40*(4).

[2] http://earth.eo.esa.int/ewarchive/cyclones/Damrey_Typhoon-sep05/

[3] http://www.theguardian.com/world/2012/jul/31/india-blackout-electricity-power-cuts

[4] http://news.sky.com/story/1414477/militant-attack-plunges-pakistan-into-darkness

[5] http://www.tripwire.com/state-of-security/latest-security-news/once-every-four-days-the-us-power-grid-is-under-attack/

[6] http://www.computerworld.com/article/2515570/network-security/siemens--stuxnet-worm-hit-industrial-systems.html

[7] Annual report 2011, The Repository for Industrial Security Incidents (RISI), Online: http://www.securityincidents.net/index.php/products/indepth/risi_annual_report/

[8] 2011 Report on Control System Cyber Security Incidents, online: http://community.controlglobal.com/content/risi-cyber-incident-report-201-calendar-year-out-risicybersecurity-pauto-automation-mfg-ma

[9] https://en.wikipedia.org/wiki/2006_European_blackout

[10] Li, M., Xu, L., Gong, H., Song, Z., Ding, L., & Liu, J. (2013, December). Study on critical lines identification. In *IEEE Proceedings of the 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)*, pp. 3132-3135.

[11] Wang, A., Luo, Y., Tu, G., & Liu, P. (2011). Vulnerability assessment scheme for power system transmission networks based on the fault chain theory. *IEEE Transactions on Power Systems*, *26*(1), 442-450.

[12] Xu, W., Jianhua, Z., Linwei, W., & Xingyang, Z. (2012, September). Power system key lines identification based on cascading failure and vulnerability evaluation. In *IEEE Proceedings of the 2012 China International Conference on Electricity Distribution*, pp. 1-4.

[13] Mohajerani, Z., Farzan, F., Jafary, M., Lu, Y., Wei, D., Kalenchits, N., ... & Skare, P. (2010, April). Cyber-related risk assessment and critical asset identification within the

power grid. In *IEEE PES Proceedings of Transmission and Distribution Conference and Exposition*, 2010, pp. 1-4.

[14] Zhu, Y., Yan, J., Tang, Y., Sun, Y. L., & He, H. (2015). Joint Substation-Transmission Line Vulnerability Assessment Against the Smart Grid. *IEEE Transactions on Information Forensics and Security*, *10*(5), 1010-1024.

[15] Ten, C. W., Manimaran, G., & Liu, C. C. (2010). Cybersecurity for critical infrastructures: attack and defense modeling. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, *40*(4), 853-865.

[16] Liu, N., Zhang, J., Zhang, H., & Liu, W. (2010). Security assessment for communication networks of power control systems using attack graph and MCDM. *IEEE Transactions on Power Delivery*, *25*(3), 1492-1500.

[17] Vellaithurai, C., Srivastava, A., Zonouz, S., & Berthier, R. (2015). CPINDEX: Cyber-Physical Vulnerability Assessment for Power-Grid Infrastructures. *IEEE Transactions on Smart Grid*, *6*(2), 566-575.

[18] Zonouz, S., Davis, C. M., Davis, K. R., Berthier, R., Bobba, R. B., & Sanders, W. H. (2014). SOCCA: A security-oriented cyber-physical contingency analysis in power infrastructures. *IEEE Transactions on Smart Grid*, *5*(1), 3-13.

[19] Ross, K. J., Hopkinson, K. M., & Pachter, M. (2013). Using a distributed agent-based communication enabled special protection system to enhance smart grid security. *IEEE Transactions on Smart Grid*, *4*(2), 1216-1224.

[20] Zhang, G., Wang, C., Zhang, J., Yang, J., Zhang, Y., & Duan, M. (2008, April). Vulnerability assessment of bulk power grid based on complex network theory. In *Proceedings of the 3rd International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*, Nanjing, China, 2008, pp. 1554-1558.

[21] Qi, J., Mei, S., & Liu, F. (2013). Blackout model considering slow process. *IEEE Transactions on Power Systems*, *28*(3), 3274-3282.

[22] Mei, S., He, F., Zhang, X., Wu, S., & Wang, G. (2009). An improved OPA model and blackout risk assessment. *IEEE Transactions on Power Systems*, *24*(2), 814-823.

[23] Chen, J., & Thorp, J. S. (2002). A reliability study of transmission system protection via a hidden failure DC load flow model. In *Proceedings of the 5th International Conference on Power System Management and Control*, 2002, pp. 384-389.

[24] Dobson, I. (2007, June). Where is the edge for cascading failure?: challenges and opportunities for quantifying blackout risk. In *IEEE Proceedings of Power Engineering Society General Meeting*, 2007, pp. 1-8.

[25] Carreras, B. A., Lynch, V. E., Dobson, I., & Newman, D. E. (2004). Complex dynamics of blackouts in power transmission systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *14*(3), 643-652.

[26] Chen, J., Thorp, J. S., & Dobson, I. (2005). Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model. *International Journal of Electrical Power & Energy Systems*, *27*(4), 318-326.

[27] Transmission reliability evaluation for large-scale systems (TRELSS): version 6.0 User's manual, EPRI, Palo Alto, CA: 2000. 1001035

[28] Eppstein, M. J., & Hines, P. D. (2012). A "random chemistry" algorithm for identifying collections of multiple contingencies that initiate cascading failure. *IEEE Transactions on Power Systems*, *27*(3), 1698-1705.

[29] Zhu, Y., Yan, J., Tang, Y., Sun, Y. L., & He, H. (2014). Resilience analysis of power grids under the sequential attack. *IEEE Transactions on Information Forensics and Security*, *9*(12), 2340-2354.

[30] Zhu, Y., Yan, J., Tang, Y., Sun, Y., & He, H. (2014, June). The sequential attack against power grid networks. In *Proceedings of the 2014 IEEE International Conference on Communications (ICC)*, pp. 616-621.

[31] Nasiruzzaman, A. B. M., & Pota, H. R. (2011, August). Critical node identification of smart power system using complex network framework based centrality approach. In *Proceedings of IEEE North American Power Symposium (NAPS)*, 2011, pp. 1-6.

[32] Zeng, K., Wen, J., Cheng, S., Lu, E., & Wang, N. (2014, February). A critical lines identification algorithm of complex power system. In *IEEE PES Proceedings of Innovative Smart Grid Technologies Conference (ISGT)*, 2014, pp. 1-5.

[33] Van Mieghem, P. (2006). *Performance analysis of communications networks and systems*. Cambridge University Press.

[34] Koç, Y., Warnier, M., Kooij, R. E., & Brazier, F. M. (2013, April). A robustness metric for cascading failures by targeted attacks in power networks. In *Proceedings of the 2013 10th IEEE International Conference on Networking, Sensing and Control (ICNSC)*, pp. 48-53.

[35] Koç, Y., Warnier, M., Kooij, R. E., & Brazier, F. M. (2013). An entropy-based metric to quantify the robustness of power grids against cascading failures. *Safety Science*, *59*, 126-134.

[36] Guo, J., Fu, Y., Li, Z., & Shahidehpour, M. (2009). Direct calculation of line outage distribution factors. *IEEE Transactions on Power Systems*, *24*(3), 1633-1634.

[37] Cuzzocrea, A., Papadimitriou, A., Katsaros, D., & Manolopoulos, Y. (2012). Edge betweenness centrality: A novel algorithm for QoS-based topology control over

wireless sensor networks. *Journal of Network and Computer Applications, 35*(4), 1210-1217.

[38] Koç, Y., Verma, T., Araujo, N., & Warnier, M. (2013, November). Matcasc: A tool to analyse cascading line outages in power grids. In *Proceedings of the 2013 IEEE International Workshop on Intelligent Energy Systems (IWIES)*, pp. 143-148.

[39] Zimmerman, R. D., Murillo-Sánchez, C. E., & Thomas, R. J. (2011). MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, *26*(1), 12-19.

[40] Mosegaard, K., & Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *J. geophys. Res*, *100*(B7), 12431-12447.

[41] Helton, J. C., & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, *81*(1), 23-69.

[42] Kauffman, S. (1995). *At Home in the Universe: The Search for the Laws of Self-organization and Complexity*. Oxford University Press, USA.

[43] Nowicki, E., & Smutnicki, C. (1996). A fast taboo search algorithm for the job shop problem. *Management Science*, *42*(6), 797-813.

[44] Zhu, Y., Yan, J., Sun, Y., & He, H. (2014). Revealing cascading failure vulnerability in power grids using risk-graph. *IEEE Transactions on Parallel and Distributed Systems*, *25*(12), 3274-3284.

[45] Bai, Q. (2010). Analysis of particle swarm optimization algorithm. *Computer and Information Science*, *3*(1), p. 180.