Cascading Failure Pattern Identification in Power Systems Based on Sequential Pattern Mining

Lu Liu, *Student Member, IEEE,* Hao Wu, *Member, IEEE,* Linzhi Li, *Student Member, IEEE,* Danfeng Shen, *Student Member, IEEE,* Feng Qian, and Junlei Liu

***Abstract*—Cascading failure simulation data contain many fault chains (FCs), which can present cascading failure propagation paths. In cascading failure process, some component outages play dominating roles in propagation, indicating the common characteristics among different FCs. The commonness is embodied as combinations of components which are vulnerable to trip and cause serious blackout consequences. In addition, a component outage will increase the outage probability of relevant components and induce dependent outage in subsequent stage. Such relevance between two components can be called component outage causality. A combination of sequential component outages with outage causalities, which exists in different FCs and leads to system load loss, can be regarded as a cascading failure pattern (CFP). Statistical characteristics of FCs indicate that CFPs are variously distributed in different FCs, can present propagation paths and cause different impacts on system blackouts. This paper proposes a cascading failure pattern (CFP) identification method based on sequential pattern mining approach. The proposed method focuses on mining CFPs from massive FCs, quantifies the influence of CFP on system blackout and identifies the critical ones. The proposed method is verified with FCs data on IEEE 39-bus and 118-bus test systems.**

***Index Terms*—Cascading failure pattern (CFP), sequential pattern mining, component outage causality, fault chain (FC).**

NOMENCLATURE

Symbols

L	An FC
P	A CFP
S	A fault state
s	A sequence, k-sequence $\|s\|$=k
N	Number of FCs
m	Number of system components
l_i	Numbered system component
$\tilde{l}_i$	Component in CFP
Ω	A set of FCs
ϑ	A set of CFPs
ξ	A set of fault states
τ	A set of optional components
α	Representation of a component
β	Representation of a sequence
C	Children sequence
F	Parent sequence
λ_s	Support of sequence s
λ_{th}	User-specified support threshold value
J	Similarity index
J_{th}	User-defined similarity threshold value
$\omega(\cdot)$	Influence weight index
$r(\cdot)$	Risk index
$D(\cdot)$	Accumulative load loss
$\eta(\cdot)$	Loss cost
$g_i^{(P)}$	Component gap between $\tilde{l}_i$ and $\tilde{l}_{i+1}$ in an FC
$t^{(P)}$	Stage of first component of P tripped in an FC
$t_0^{(P)}$	Stage of final component of P tripped in an FC
$\|\cdot\|$	Number of elements in a set, or length of a sequence
$\langle\cdot\rangle$	Component combination with specific order
$(\cdot)$	Component combination with diverse orders
$\rightarrow$	Causality between component outages

Concepts

sequence	A series of sequential component outages
subsequence/ supersequence	s_1 ($\|s_1\|$=k_1) is a subsequence of s_2 ($\|s_2\|$=k_2) (s_2 is a supersequence of s_1). ($\exists 1 \le j_1 < j_2 < \cdots < j_{k_1} \le k_2, l_1^{(1)}$=$l_{j_1}^{(2)}, \cdots, l_{k_1}^{(1)}$=$l_{j_{k_1}}^{(2)}$)
s_2 *contains* s_1	s_1 is a subsequence of s_2 (s_2 is a supersequence of s_1)
support	Number of sequences containing the target sequence
frequent sequence	A sequence with support greater than λ_{th}

I. INTRODUCTION

WITH the rapid growth of power demands and the updating of power grid structure, cascading failures occur frequently in recent decades, which severely influence social life and economic safety. There are several large-scale cascading failure blackouts worldwide, for example, 2003 Northeast blackout over U.S. and Canada [1], 2006 European blackout [2], 2012 India blackouts [3] and 2019 Great Britain blackout [4]. A cascading failure is triggered by random initial contingencies such as natural disasters or component defects, and followed by subsequent propagation of dependent component outages such as transformer or transmission line outages [5], [6]. It is significant to analyze cascading failure propagation so that factors causing severe blackout consequences can be better identified and mitigated.

At present, cascading failure simulations are the main methods for cascading failure analysis in power systems. Among these simulation methods, the dynamic simulation models (OPA Model [7], [8], Manchester Model [9], Cosmic

L. Liu, H. Wu, L. Li and D. Shen are with the College of Electrical Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China (e-mail: liuluzjuee@zju.edu.cn; zjuwuhao@zju.edu.cn; lilinzhilee@gmail.com; dfshen@zju.edu.cn). *(Corresponding author: Hao Wu.)*

F. Qian and J. Liu are with the Guangdong Power Grid Power Dispatching Control Center, Guangzhou 510000, China (e-mail: 25326271@qq.com; 18802000394@163.com).

[10], etc.) reflect component dynamic mechanisms with physical details, but are time-consuming; the high-level statistical models (Cascade [11], Branching Process [12]) can statically simulate cascading failure, but can not retain physical details about power systems and also can not describe in detail the component outage propagation process.

Because the process of cascading failure contains valuable information, such as topological feature, power flow distribution, component outage relevance and system load loss, the data-based analysis methods based on historical or simulated cascading failure data are receiving growing attention recently for analyzing cascading failure propagation. Ref. [13] and [14] identified critical components in cascading failure considering topological feature and power flow distribution. Ref. [13], [15]–[17] studied the relevance between two component outages by constructing interaction graph to describe the possible cascading failure propagation path. The relevance is quantified by considering propagation probability of two consecutive component outages. Ref. [18] and [19] studied the common characteristics in propagation paths based on occurrence frequency, and describe the commonness as a series of consecutive component outages. Above researches indicate that there are indeed critical components and outage relevance in cascading failure, which are the dominating factors in propagation. In fact, since power flow distribution and component connections are changeable in cascading failure process, criticalness of component and outage relevance are related to fault state. Furthermore, the dominating role in propagation should consider occurrence frequency and blackout consequence simultaneously. It is a combination of sequential component outages that may frequently develop into serious fault states and result in severe blackouts.

Therefore, we focuses on mining component combinations containing critical components and outage relevance. If such component combination exists in many propagation paths with high occurrence frequency and severe consequence, namely dominating the propagation paths, then it can be regarded as a cascading failure pattern independent of propagation paths. The pattern can help better understand why and how cascading failures occur and propagate, and hence is worth researching.

In this paper, we propose the concept of "cascading failure pattern (CFP)" to embody the above pattern, and propose a cascading failure pattern identification method to identify CFPs from massive cascading failure data (i.e. fault chain (FC) data). The proposed method is also a data-based method. The CFP is a combination of sequential component outages with outage causalities, playing dominating role in propagation. A noteworthy CFP, called critical CFP, has the characteristics of high occurrence frequency, severe blackout consequence and variable distributions in different FCs. Because the component outages in a CFP may be discontinued-occurred outages in different FCs, we propose a method based on sequential pattern mining approach [20], [21] to ensure CFP mining requirements. The approach is one type of classic data mining methods, which has been widely used in web protocol extraction, DNA sequence analysis, customer purchase behavior [22]–[24], etc.

The remainder of this paper is organized as follows. Section II gives data structure for cascading failure and proposes the definition of cascading failure pattern (CFP). Section III proposes the cascading failure pattern identification method to mine CFPs from massive fault chains (FCs), and quantification method to identify critical CFPs. Section IV verifies the identification results on IEEE 39-bus and 118-bus test systems, and discusses the CFP-based mitigation measure in IEEE 118-bus test system. Section V draws the conclusions.

II. Formulation of Cascading Failure Pattern

A. Data Structure for Cascading Failure

In this paper, we identify cascading failure patterns (CFP) from fault chains (FCs) [25] obtained from the cascading failure simulation data. An FC is a sequence of consecutive component outages, which represents the propagation path of cascading failure. The components refer to transformers and transmission lines.

Let FCs be $L_i, i = 1, \ldots, N$, where N is the number of FCs. The i-th FC L_i is denoted by

$$L_i = \langle l_1^{(i)}, \ldots, l_j^{(i)}, \ldots, l_n^{(i)} \rangle, \tag{1}$$

where the $l_j^{(i)}$ is the j-th tripped component in L_i; and n is the number of tripped components in L_i, i.e., the length of L_i.

Furthermore, we obtain the $fault\ state$ by decomposing FCs according to the outage sequence. In this paper, define a fault state as a sequence of consecutive component outages. The n-length FC L_i is decomposed into n fault states $S_1^{(i)}, \ldots, S_j^{(i)}, \ldots, S_n^{(i)}$, and the j-th fault state is

$$S_j^{(i)} = \langle l_1^{(i)}, \ldots, l_j^{(i)} \rangle \rightarrow D(S_j^{(i)}), \tag{2}$$

where $D(S_j^{(i)})$ is the accumulative load loss of fault state $S_j^{(i)}$ and is called the $fault\ state\ loss$. For the set of all decomposed fault states, merge the identical ones (namely fault states with the same component outage sequence), and define the obtained fault state set as ξ, which consists of distinct fault states.

In this paper, the *sequence* is a series of sequential component outages. For example, a fault state is a sequence of consecutive component outages, and a CFP is a sequence of sequential component outages (may contain component gap, introduced in Section II-B). Define the $loss\ cost$ for an arbitrary sequence s obtained from fault state set ξ as

$$\eta(s) = D(\xi_S), \tag{3}$$

where ξ_S represents the fault state S in ξ identical to sequence s; and $D(\xi_S)$ is the fault state loss of S. Equation (3) quantifies the general load loss of sequence s in cascading failure.

B. Cascading Failure Pattern

In cascading failure process, a component outage causes excessive flow stress on other relevant components due to power flow redistribution, and further induces the subsequent dependent component outage [15], which reflects the outage causality between components in FCs. In addition, in a complex power system, due to factors such as network topology,

generation and load distribution and component parameters (e.g. transmission line capacity), some critical component outages appear frequently and cause severe blackout consequences in cascading failure. Because of the causalities among critical components, they will appear in FCs in the form of component combinations, playing dominating roles in cascading failure propagation [26]. Furthermore, a component tripped in different FCs will have different influence on blackouts. It is because the outage is tripped under different fault states, and the system might need to shed different amount of loads to maintain the equilibrium of power system [18]. In summary, there exists combinations of component outages with outage causalities in FCs that can develop into serious fault states and aggravate system load losses. If such component combination exists in many FCs with high occurrence frequency and severe consequence, then it can be regarded as a cascading failure pattern independent of FCs.

Therefore, this paper proposes the concept of "cascading failure pattern (CFP)" to describe above combination of component outages with outage causalities. CFPs can be mined from FCs and is a subsequence of FCs. It contains the information of outage relevance and outage propagation, which is helpful in illustrating cascading failure propagation.

Define CFP as the minimal sequence of sequential component outages that results in system load loss. By "minimal", we mean that no subsequences of the CFP will result in system load loss. The CFP is denoted by

$$P = \langle \tilde{l}_1 \rightarrow \cdots \rightarrow \tilde{l}_j \rightarrow \tilde{l}_{j+1} \rightarrow \cdots \rightarrow \tilde{l}_k \rangle, \tag{4}$$

where $\tilde{l}_j$ is the j-th component in P; the symbol $\sim$ distinguishes the component in CFP from that in FC; k is the length of CFP; and $\rightarrow$ represents component outage causality between two adjacent components $\tilde{l}_j$ and $\tilde{l}_{j+1}$ in P.

Furthermore, Define $\eta(P)$ as the loss cost of P according to equation (3), since CFP is a sequence of sequential component outages that can aggravate system load losses. Define $t^{(P)}$ as *outage stage*, which is the stage that the first component $\tilde{l}_1$ of P tripped in a certain FC. Define $g_j^{(P)}$ as *component gap* between two adjacent components $\tilde{l}_j$ and $\tilde{l}_{j+1}$ in P, which is the number of in-between components of $\tilde{l}_j$ and $\tilde{l}_{j+1}$ in a certain FC. Obviously, if $g_j^{(P)} = 0$, $\tilde{l}_j$ and $\tilde{l}_{j+1}$ are tripped one after another consecutively in an FC; otherwise, $\tilde{l}_j$ and $\tilde{l}_{j+1}$ are tripped discretely in an FC. For example, for an FC $L_1 = \langle l_1^{(1)}, \cdots, l_j^{(1)}, \cdots, l_n^{(1)} \rangle$ which contains a CFP $P_1 = \langle \tilde{l}_1^{(1)} \rightarrow \tilde{l}_2^{(1)} \rangle$, and a fault state $S_2 = \langle l_1^{(2)}, l_2^{(2)} \rangle$. Loss cost of CFP P equals to load loss of fault state S_2, denoted by $\eta(P) = D(S_2)$, if conditions meet $\tilde{l}_1 = l_1^{(2)}$ and $\tilde{l}_2 = l_2^{(2)}$. Outage stage $t^{(P)} = j$, if conditions meet $\tilde{l}_1 = l_j^{(i)}$. Component gap $g_1^{(P)} = n - j$, if conditions meet $\tilde{l}_1 = l_j^{(i)}, \tilde{l}_2 = l_n^{(i)}$.

Statistical characteristics of FCs indicate that CFPs are variously distributed in different FCs. For better understanding the structure of CFPs in FCs, we use a case in Fig.1. Assume there are totally 3 CFPs P_1, P_2, P_3 in 4 FCs L_a, L_b, L_c, L_d, where $P_1 = \langle \tilde{l}_1^{(1)} \rightarrow \tilde{l}_2^{(1)} \rangle$, $P_2 = \langle \tilde{l}_1^{(2)} \rightarrow \tilde{l}_2^{(2)} \rightarrow \tilde{l}_3^{(2)} \rangle$, $P_3 = \langle \tilde{l}_1^{(3)} \rightarrow \tilde{l}_2^{(3)} \rightarrow \tilde{l}_3^{(3)} \rangle$; and $L_a = \langle l_1^{(a)}, \ldots, l_{12}^{(a)} \rangle$, $L_b = \langle l_1^{(b)}, \ldots, l_{10}^{(b)} \rangle$, $L_c = \langle l_1^{(c)}, \ldots, l_5^{(c)} \rangle$, $L_d = \langle l_1^{(d)}, \ldots, l_8^{(d)} \rangle$. As

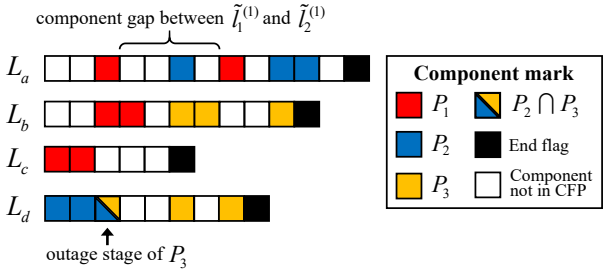

Fig. 1. An example of CFP structure in FCs. A block chain corresponds to an FC. A block corresponds to a component outage, the order of blocks corresponds to the sequence of component outages, and the number of blocks of a block chain corresponds to the length of an FC. The colored blocks represent components in CFPs, the same colored blocks represent the components in the same CFP, different colors distinguish different CFPs. The intersection block ($P_2 \cap P_3$) represents the common component that different CFPs (P_2 and P_3) share.

shown in Fig.1, the outage stage and component gap of a CFP are various in different FCs. For example, $t^{(P_1)} = 3$ in L_a and L_b, $t^{(P_1)} = 1$ in L_c; $g_1^{(P_1)} = 4$ in L_a, $g_1^{(P_1)} = 0$ in L_b and L_c. P_2 and P_3 can draw the similar conclusions. Furthermore, an FC may contain more than one CFP. The occurrence of multiple CFPs in an FC can be summarized into three types. (i) Sequential: CFPs are induced sequentially, like P_1 and P_3 in L_b. (ii) Alternate: CFPs occur alternatively. The outages of a CFP can be induced in the component gaps of other CFP, like $\tilde{l}_1^{(2)}$ of P_2 in gap $g_1^{(P_1)}$ of P_1 in L_a. (iii) Overlapped: CFPs share the common component outages in a certain FC, which means component outages of a CFP can induce another CFP, like P_2 and P_3 in L_d have $\tilde{l}_3^{(2)} = \tilde{l}_1^{(3)} = l_3^{(d)}$.

In addition, there exists scenarios that one or none CFP is available in an FC. For an FC containing one CFP, it means the components, tripped after all of components in the CFP have been tripped, will increase the length of FC but no longer contribute to aggravate system load loss, like P_1 in L_c. For an FC containing no CFP, it means the cascading failure process will not cause system load loss. It is relatively safe, since no blackout happen and cascade process can be terminated by relay protection.

In conclusion, the CFP that we project to identify from massive FCs has the following characteristics.

1) Outage stage t is variable, the first component of a CFP can be initial outage, or be induced by other outages in subsequent stages.
2) Component outage causality is contained in the CFP, which represents the relevance between two adjacent components of the CFP with variable component gap g. The components of CFPs are either tripped one after another consecutively ($g = 0$) or one after another discretely ($g \neq 0$) in a certain FC.
3) One or more than one CFP can appear in an FC, and there are three CFP occurrence types, sequential, alternate and overlapped.
4) The CFP is vulnerable to be tripped and will aggravate system load loss, playing dominating role in cascading failure propagation.

Obviously, above characteristics make it complex and difficult for CFP identification and separation from massive FCs.

To overcome the difficulties, we propose a cascading failure pattern identification method to effectively identify CFPs.

III. Cascading Failure Pattern Identification Method

A. Cascading Failure Pattern Mining

Since the cascading failure can be regarded as Markovian process, the randomness of component outages causes diverse component outage sequences [27], namely the fault chains (FCs). Different FCs contains different CFPs, which lead to different system blackout consequences.

Considering CFP characteristics in Section II-B, this part proposes a cascading failure pattern mining algorithm to realize CFP mining. The proposed method is based on sequential pattern mining approach, which is used for finding statistically relevant and frequent patterns hidden in data sequences [20].

The proposed method aims to mine CFPs from fault states by constructing search trees. The search tree is organized by depth and at each depth there exist some sequences with the same length. Let F_i denote the set of sequences at depth i, the sequence in which is called parent sequence $F_j^{(i)}, j = 1, \ldots, |F_i|$. Define child sequence as a subsequence of a parent sequence after removing an arbitrary component. Let $C_{F_j^{(i)}}$ denote the child sequence set of parent sequence $F_j^{(i)}$. Then, the child sequence set at depth $i+1$ is the union of all these child sequence sets, namely

$$C_{i+1} = \bigcup_{j=1}^{|F_i|} C_{F_j^{(i)}}. \tag{5}$$

The search tree starts from a given fault state S, which makes up the 0 depth parent sequence set $F_0 = \{S\}$, and iteratively stretches by generating the next-depth parent sequence set F_{i+1} from the current parent sequence set F_i. The next-depth parent sequences F_{i+1} are selected by searching fault states with non-zero loss cost in ξ, namely

$$F_{i+1} = \{\beta | \beta \in \xi \cap C_{i+1}, \eta(\beta) \neq 0\}. \tag{6}$$

Repeat the above process until there is no new parent sequence generated, and then the algorithm is terminated. The mined CFPs are the parent sequences whose children sequences are all of zero loss cost value. It should be noted that, the above CFPs can include single-component parent sequences, since the child sequence of them is emptyset with zero loss cost value. The pseudocode of this algorithm is in Algorithm 1.

An application example of Algorithm 1 is illustrated in Fig.2. Take $\langle A, B, C, D, E\rangle$ as the given fault state, and the mined CFPs are $\langle A, D, E\rangle, \langle B, D\rangle$.

To realize the mining of all CFPs, we need to conduct above algorithm for all fault states in ξ and generate multiple search trees respectively. If a parent sequence of the ongoing search tree has already appeared in an existing search tree, subsequent stretching of parent sequence in the ongoing search tree will be merged with existing one, for avoiding unnecessary calculations.

Algorithm 1 Cascading Failure Pattern Mining Algorithm

Input: Given fault state S, fault state set ξ.

1: Parent sequence at depth 0: $F_0 = \{S\}$.
2: **for** $i = 0; F_i \neq \emptyset; i = i + 1$ **do**
3: **for** $j = 1; j \leq |F_i|; j = j + 1$ **do**
4: Generate child sequence set $C_{F_j^{(i)}}$ of the j-th parent sequence $F_j^{(i)}$ in F_i.
5: **end for**
6: Calculate child sequence set C_{i+1} by (5).
7: Calculate parent sequence set F_{i+1} by (6).
8: **end for**
9: Calculate the set of all parent sequences: $F = \bigcup_i F_i$
10: Obtain the set of mined CFPs: $\{\beta | \beta \in F, C_\beta \notin F\}$, where C_β denotes the child sequence set of β.

Output: The CFP set.

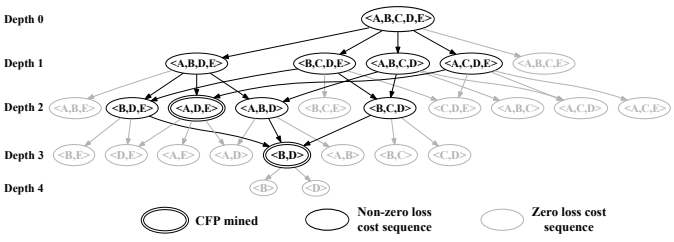

Fig. 2. An Example of cascading failure pattern mining.

B. Common CFP Mining

Different CFPs may have the same loss cost, since some components may be functionally similar in propagation [13]. For example, a set of CFPs mined by Algorithm 1, $\langle l_8 \to l_{22} \to l_{37}\rangle, \langle l_{116} \to l_8 \to l_{37}\rangle, \langle l_8 \to l_{98} \to l_{97} \to l_{37}\rangle$, share the same loss cost and have common components l_8 and l_{37}. The common subsequence $\langle l_8 \to l_{37}\rangle$ can be mined to represent the above set, which is regarded as the *common CFP*.

Define the common CFP as a common subsequence of a set of CFPs with the same load cost, which contains common components with outage causalities and appears frequently in the CFP set. Weakening the propagation of common CFP can reduce the occurrence frequency of the set of CFPs at the same time. Therefore, in order to reserve and highlight the common components with outage causalities in different CFPs, we use common CFP to represent a set of CFPs with the same loss cost. For convenience, all CFPs mined by Algorithm 1 are called original CFPs, and all common CFPs mined by Algorithm 2 are called final CFPs in the following.

First, conduct CFP classification to classify original CFPs with the same loss cost into multiple CFP sets, making sure that information of critical component and outage causalities will not be drown out. The CFPs in a CFP set ϑ satisfies

$$J(P_i^{(\vartheta)}, P_j^{(\vartheta)}) = \frac{|P_i^{(\vartheta)} \cap P_j^{(\vartheta)}|}{|P_i^{(\vartheta)} \cup P_j^{(\vartheta)}|} \geq J_{\text{th}}, \tag{7}$$

where $P_i^{(\vartheta)}, P_j^{(\vartheta)} \in \vartheta, i, j \in \{1, 2, \cdots, |\vartheta|\}$ and $i \neq j$; and J_{th} is the support threshold value.

Second, mine common CFP from the CFP set ϑ using common CFP mining algorithm, which is based on sequential pattern mining approach [20], [21]. For a CFP set ϑ, we obtain

the maximum frequent sequences by constructing search trees. By "maximum", the obtained sequence is ensured not to be contained by other subsequences of CFPs in ϑ. Similar to Algorithm 1, F_i denotes the parent sequence set at depth i, whose parent sequence is denoted by $F_j^{(i)}, j = 1, \ldots, |F_i|$. Define child sequence as a supersequence of parent sequence after adding a new component from optional component set τ. Let $C_{F_j^{(i)}}$ denotes child sequence set of parent sequence $F_j^{(i)}$, namely

$$C_{F_j^{(i)}} = \{\beta' | \beta' = \langle \beta \rightarrow \alpha \rangle, \beta \in F_j^{(i)}, \alpha \in \tau_i\}. \quad (8)$$

The search tree starts from empty set at depth 0 (i.e. $F_0 = \{\}$), and iteratively stretches by generating the next-depth parent sequence set F_{i+1} from the current parent sequence set F_i. The next-depth parent sequences F_{i+1} are selected by support threshold λ_{th}, namely

$$F_{i+1} = \{\beta | \beta \in C_{i+1}, \lambda_\beta \geq \lambda_{th}\}, \quad (9)$$

where λ_β is the support of sequence β among CFP set ϑ which is calculated using the support calculation algorithm in section III-C; and λ_{th} is similarity threshold value.

Repeat above process until there is no new parent sequence generated. Then, if there exists only one parent sequence at the final depth, the common CFP is the very parent sequence; if there exists more than one parent sequence at the final depth, the common CFP is the intersection of parent sequences. The sequences to be intersected may simultaneously exist different causalities between two components (e.g. $l_i \rightarrow l_j$ and $l_j \rightarrow l_i$). We use $(\cdot)$ to describe the component combination containing diverse orders, and $\langle \cdot \rangle$ describe the component combination with specific order. The pseudocode is given in Algorithm 2.

Algorithm 2 Common CFP Mining

Input: A CFP set ϑ.

1: Parent sequence at depth 0: $F_0 = \{\}$.
2: Optional component set at depth 0: $\tau_0 = \{\alpha | \alpha \in \{l_1, \ldots, l_m\} \cap \vartheta\}$, where $\{l_1, \ldots, l_m\}$ are the numbered components in the system.
3: Child sequence at depth 1: $C_1 = \tau_0$.
4: Generate parent sequence F_1 by (9).
5: **for** $(i = 1; F_i \neq \emptyset; i = i + 1)$ **do**
6: **for** $(j = 1; j \leq |F_i|; j = j + 1)$ **do**
7: Generate child sequence set $C_{F_j^{(i)}}$ of the i-th parent sequence $F_j^{(i)}$ by (8).
8: **end for**
9: Calculate child sequence set C_{i+1} by (5).
10: Calculate parent sequence set F_{i+1} by (9).
11: Calculate optional component set at depth $i+1$: $\tau_{i+1} = \{\alpha | \alpha \in \{l_1, \ldots, l_m\} \cap F_{i+1}\}$.
12: **end for**
13: Obtain the set of maximum frequent sequences: $\{\beta | \beta \in F_{i-1}, F_i = \emptyset\}$.
14: Obtain the common CFP: $F = \bigcap_j^{|\beta|} \beta_j$.

Output: The common CFP.

C. Support Calculation

The proposed support calculation algorithm aims to calculate how many sequences in a given sequence set contain the target sequence (e.g. the target common CFP in a CFP set ϑ).

First, construct the position matrix $\nu \in \mathbb{R}^{m \times n}$, using total m components $l_i, i = 1, \ldots, m$ and the given sequence set with n sequences $\{s_j | j = 1, \ldots, n\}$. An arbitrary entry ν_{ij} of ν denotes the position value of the i-th components in the j-th sequence $s_j = \langle l_1^{(j)} \cdots l_k^{(j)} \rangle$, namely

$$\nu_{ij} = \begin{cases} r, & \exists 1 \leq r \leq k, l_r^{(j)} = l_i \\ 0, & \text{else} \end{cases}, \quad (10)$$

where $\nu_{ij} = 0$ signifies l_i does not appear in s_j. For example, considering sequence $s_j = \langle l_1^{(j)}, l_2^{(j)}, l_3^{(j)} \rangle = \langle l_{u_1}, l_{u_2}, l_{u_3} \rangle$ where $1 \leq u_1, u_2, u_3 \leq m$, the j-th column of ν will be $\nu_{u_1 j} = 1, \nu_{u_2 j} = 2, \nu_{u_3 j} = 3$, and $\nu_{ij} = 0, \forall i \notin \{u_1, u_2, u_3\}$.

Then, calculate the support of target sequence $s_0 = \langle l_1^{(0)}, \ldots, l_i^{(0)}, l_{i+1}^{(0)} \ldots, l_k^{(0)} \rangle$ among the given sequence set $\{s_j | j = 1, \ldots, n\}$. Let $T_i \in \mathbb{R}^{1 \times n}, i = 1, \ldots, k$ denote a set of position values that component $l_i^{(0)}$ located in the given sequence set $\{s_j | j = 1, \ldots, n\}$, namely, the elements of the row (corresponding to $l_i^{(0)}$) of position matrix ν. According to the definition of 'subsequence', s_j containing s_0 means component $l_i^{(0)}$ must appear before $l_{i+1}^{(0)}$ in s_j the same as they are in s_0. Mathematically, this feature means $T_{i,j} \leq T_{i+1,j}, i = 1, \ldots, k$, where $T_{i,j}$ is the j-th element of T_i and denotes the position value of $l_i^{(0)}$ in s_j. Therefore, the support of s_0, denoted by λ_{s_0}, is calculated by

$$\lambda_{s_0} = |\{j | 0 < T_{1,j} < \cdots < T_{k,j}, j \in \{1, \ldots, n\}\}|. \quad (11)$$

D. Critical CFP Identification

This part identifies critical CFPs from mined final CFPs which are obtained after Algorithm 1 and 2. The critical CFP is a CFP that has high occurrence frequency and severe impacts on system blackout consequences.

First, considering the conditions that variable component gap and outage stage of a CFP in different FCs and multiple CFPs in an FC, the average impact of a CFP on FCs is quantified by influence weight index ω, which is denoted by

$$\omega_P = \frac{1}{\lambda_P} \sum_{i=1}^{\lambda_P} \frac{\eta(P)}{D(S_{t_0^{(P)}}^{(i)})}, \quad (12)$$

where λ_P is the support of CFP P among the given N FCs, calculated using support calculation algorithm in section III-C; $t_0^{(P)}$ is the stage that final component of P tripped in an FC; $\eta(P)$ is the loss cost of P; and $D(S_{t_0^{(P)}}^{(i)})$ is the fault state loss of fault state $S_{t_0^{(P)}}^{(i)} = \langle l_1^{(i)}, \ldots, l_{t_0^{(P)}}^{(i)} \rangle \subseteq L_i, L_i \in \Omega_P$ and Ω_P is a set of FCs that contains P, $|\Omega_P| = \lambda_P$. The CFP located more frequently in the earlier stage is assigned a greater value of influence weight index $\omega \in (0, 1]$, which emphasizes the significance of the CFP.

Then, the risk of a CFP on system blackouts is quantified

by pattern risk index (PRI)

$$r(P) = \frac{\lambda_P}{N}\eta(P)w_P, \tag{13}$$

where N is the number of FCs; and ω_P is the influence weight index calculated by (12). PRI evaluates the risk degree of CFP on system load loss, and the higher value refers to the more severe consequence of the CFP that will influence cascading failure propagation, namely, more critical in cascading failure.

Furthermore, the risk of a component on system blackouts can be further calculated by component risk index (CRI)

$$r(l_i) = \sum_{j\in|\vartheta_{l_i}|} r(P_j)\frac{\lambda_{P_j}}{\lambda_{\vartheta_{l_i}}}, \tag{14}$$

where ϑ_{l_i} is a set of CFPs that contains component l_i; $P_j \in \vartheta_{l_i}, j = \{1,\ldots,|\vartheta_{l_i}|\}$; λ_{P_j} is the support of P_j among all FCs; and $\lambda_{\vartheta_{l_i}} = \sum_{j\in|\vartheta_{l_i}|}\lambda_{P_j}$ is the sum of support of P_j in ϑ_{l_i}. CRI quantifies the synthetic impact of component on cascading failure propagation. The high value indicates that the component may occur more frequently in CFPs and is more critical in propagation, the outage of which can cause more severe consequence on system blackout.

IV. Simulation Results

This section presents results for the proposed cascading failure pattern identification method. The feasibility and effectiveness of the proposed method are validated in IEEE 39-bus test system. The effectiveness of critical CFPs identification and mitigation are validated in IEEE 118-bus test system. The two test systems are adopted from [28]. DC power flow model [7] is adopted in this paper for cascading failure simulation and generating N FCs. It should be noted that DC power flow model is only one model that can be used to generate FCs needed by the proposed method. Other models (e.g. AC power flow model) are also applicable to our proposed method.

Both power flow model and the proposed method are implemented with MATLAB and all tests are carried out on a workstation with Intel Xeon E5-2660v4+64GB RAM. The similarity threshold and support threshold are set to $J_{\text{th}} = 0.45$ and $\lambda_{\text{th}} = 2$ respectively.

A. The IEEE 39-bus test system

The IEEE 39-bus test system includes 39 buses (with 10 generators), 46 components (including 34 transmission lines and 12 transformers). Cascading failure is triggered by initial random N-2 contingencies and generates 100,000 FCs in total.

1) Critical CFP Identification: The proposed method totally identifies 50 final CFPs, involving 39 components. For better demonstrating the composition of mined CFPs, 14 final CFPs with PRI greater than 2% of the maximum PRI are presented in Fig.3. It shows that different CFPs can contain different numbers of components, and the components in different CFPs have inequable impacts on blackouts. A low-risk CFP may contain high-risk components, and similarly, the combination of low-risk components may represent a high-risk CFP, which signifies the identification of critical CFP is meaningful and different from critical component identification.

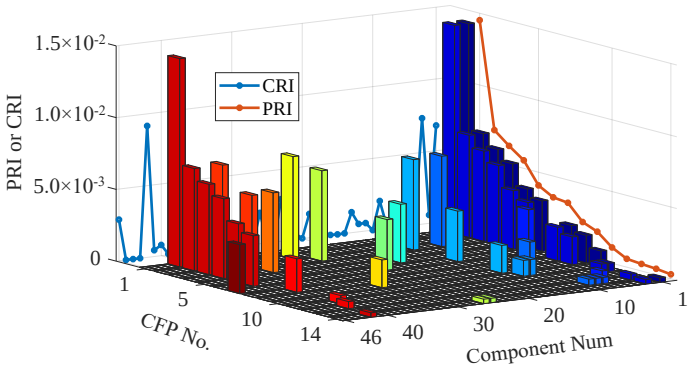

Fig. 3. Composition of CFPs in IEEE 39-bus test system. For 3-D bars, a bar represents one specific component in a certain CFP, the height of bar is the corresponding PRI of the CFP, and the bar colors distinguish different components. For two 2-D curves, blue curve presents CRI of 46 components, and red curve presents PRI of 14 critical CFPs.

TABLE I
Critical CFPs with high PRI in IEEE 39-bus test system

No.	CFP	PRI	FC Prop.
1	$\langle(l_3, l_{42}) \to l_1\rangle$	1.45×10^{-2}	40.711%
2	$\langle(l_{28}, l_{38}) \to l_3 \to l_{42} \to l_1\rangle$	7.13×10^{-3}	3.612%
3	$\langle(l_9, l_{13}, l_{26}) \to l_3 \to l_{42} \to l_1\rangle$	6.30×10^{-3}	2.668%
4	$\langle(l_{35}, l_{38}) \to l_3 \to l_{42} \to l_1\rangle$	5.56×10^{-3}	2.459%
5	$\langle l_{19} \to l_3 \to l_{42} \to l_1\rangle$	4.08×10^{-3}	2.210%
6	$\langle(l_{13}, l_{23}) \to l_3 \to l_{42} \to l_1\rangle$	3.52×10^{-3}	1.963%
7	$\langle l_{46} \to l_5\rangle$	3.41×10^{-3}	2.861%
8	(l_1, l_3, l_{40})	2.34×10^{-3}	8.406%
9	$\langle l_{30} \to l_3 \to l_9 \to l_{13} \to l_1\rangle$	1.93×10^{-3}	1.002%
10	$\langle(l_{11}, l_{12}) \to l_1\rangle$	1.10×10^{-3}	0.800%
11	(l_2, l_3, l_{40})	5.81×10^{-4}	2.249%
12	$\langle l_{41} \to l_5\rangle$	4.94×10^{-4}	1.531%
13	$\langle l_8 \to l_3 \to l_9 \to l_7\rangle$	4.26×10^{-4}	0.481%
14	$\langle(l_{25}, l_{26}) \to l_3 \to l_{42} \to l_1\rangle$	3.09×10^{-4}	0.205%

The identified first 14 final CFPs with high PRI are defined as critical CFPs, as shown in Table I, where the value of 'FC Prop.' column represents the proportion of FCs containing the corresponding critical CFP shown in the 'CFP' column; and $(\cdot)$ represents the components whose outage sequences are diverse. For example, the No.1 CFP P_1 $\langle(l_3, l_{42}) \to l_1\rangle$ indicates there are 2 CFPs in the actual propagation paths, namely $\langle l_3 \to l_{42} \to l_1\rangle$ and $\langle l_{42} \to l_3 \to l_1\rangle$. Table I shows that: (i) CFPs with the same components and different orders have similar impact on cascading failure propagation, like the two CFP $\langle l_3 \to l_{42} \to l_1\rangle$ and $\langle l_{42} \to l_3 \to l_1\rangle$ in P_1. (ii) Different critical CFPs might contain the same outage causalities, like $P_1, P_2, P_3, P_4, P_5, P_6, P_{14}$ all have causalities $l_3 \to l_{42}$ and $l_{42} \to l_1$. The above critical CFPs belong to different types of CFPs because the loss costs of which are different from each other. (iii) A CFP with high occurrence frequency do not necessary carry higher risks, which signifies the importance of critical CFP identification. For example, the occurrence frequency of $P_8 = (l_1, l_3, l_{40})$ ranks 2nd among critical CFPs (i.e. 8.406% proportion of FCs contain P_8), but the PRI ranking of P_8 is relatively lower. (iv) P_1 has relatively much higher occurrence frequency than others and 40.711%

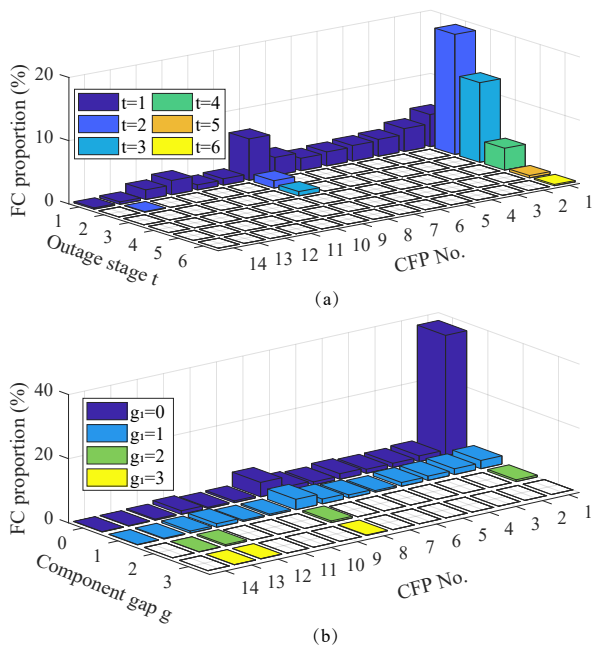

Fig. 4. Outage stage t and component gap g of critical CFPs in IEEE 39-bus test system. (a) Different t value of a CFP in FCs. A bar represents t of the CFP equals to a certain value in an FC, the height of bar is the corresponding FC proportion, and the bar colors distinguish different values of t. (b) Different g value of a CFP in FCs. g_1 represents the 1-th gap of a CFP. A bar represents g_1 of the CFP equals to a certain value in an FC, the height of bar is the corresponding FC proportion, and the bar colors distinguish different values of g_1.

FCs contains P_1. It indicates that P_1 is more vulnerable to be tripped and prone to aggravate system, which plays dominating role in propagation.

2) Outage stage and component gap: The occurrence frequency distribution of outage stage t (i.e. stage of the first component of CFP tripped in an FC) and component gap g (i.e. number of in-between components of two adjacent components of CFP in an FC) of critical CFPs in different FCs are shown in Fig.4.

Fig.4(a) shows that outage stage of a CFP is various in different FCs. (i) The CFPs are mainly triggered in initial stage. For example, $t = 1$ is the main outage stage value for most CFPs (e.g. $P_2, P_3, \ldots, P_{14}$). It means that it is more likely for the first component of a CFP to trip in the initial stage and induce the rest components of the CFP to be tripped sequentially in the subsequent stages. (ii) The CFPs can also be induced by other outages in the subsequent stages. For example, P_1, P_8 and P_{13} have stages $t^{(P_1)} = \{1, 2, 3, 4, 5, 6\}$, $t^{(P_8)} = \{1, 2, 3\}$, and $t^{(P_{13})} = \{1, 2\}$ respectively. In addition, P_1 has relatively higher occurrence frequency in stages $t^{(P_1)} = 2$ and 3, which means that P_1 is vulnerable to be induced by diverse outages in the subsequent stages. Fig.4(a) indicates that if critical CFP can be mitigated in the early stages, the occurrence frequency of critical CFP can be effectively reduced.

Fig.4(b) shows that the component gap of a CFP is various in different FCs. Since there is a gap between arbitrary two adjacent components of a CFP, more than one gap will be contained in a CFP when the number of components in the CFP is greater than 2. In order to present the different g value of a CFP in FCs, we take the 1-th gap of a CFP g_1 as an example in Fig.4(b). Other gaps of a CFP can obtain similar observation results.

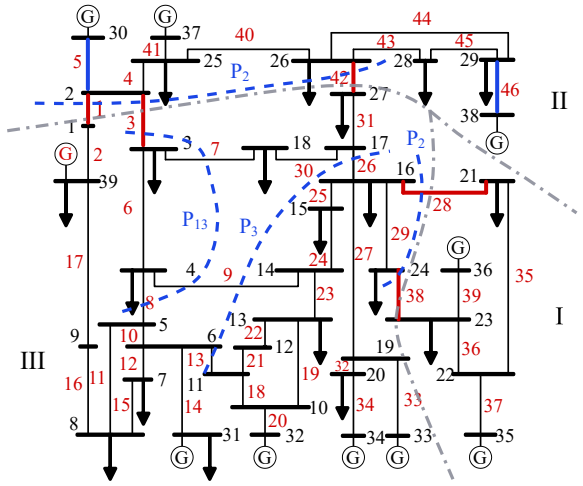

Fig. 5. Topology diagram of IEEE 39-bus test system. The gray dot-dash lines isolate system into 3 partitions according to P_2 and red bold line represents components in P_2. The blue dash lines link components of transmission channels in P_2, P_3, P_{13} respectively. The blue bold lines represent components related to generators in P_7.

The various g_1 in Fig.4(b) validates that the component outage of a CFP can dependently induce the component with outage causality in the CFP to be tripped, whether consecutively($g = 0$) or discretely($g \neq 0$). For example, $g = 0$ and $g = 1$ are the main component gap values for all the CFPs. Besides, some CFPs have gaps with greater values (e.g. P_8, P_{12} and P_{13} all have $g_1 = \{0, 1, 2, 3\}$), though with relatively lower occurrence frequency (i.e. lower FC proportions). The greater value of gap might cause the blackouts with longer propagation paths and more severe system accumulative load losses. Fig.4(b) indicates that if the propagation of critical CFP can be mitigated according to outage causalities, the occurrence frequency of critical CFP and large-scale blackouts can be effectively reduced.

3) Topological Characteristic: Critical CFP in Table I has clear topology characteristics. To view this, some critical CFPs are presented in Fig. 5, and conclusions are drawn in the following.

(i) Critical CFPs can clearly isolate the network into multiple sub-networks. For example, the sequential outages of components in $P_2 = \langle (l_{28}, l_{38}) \rightarrow l_3 \rightarrow l_{42} \rightarrow l_1 \rangle$ can isolates the system into 3 partitions gradually (i.e. partition I, II and III in Fig.5). The disconnection between sending-end partitions I,II and receiving-end partition III will further result in the load shedding in partition III for the equilibrium of power and load.

(ii) Critical CFPs reveal important transmission channels of the network, e.g. combinations $\{l_{28}, l_{38}\}$ and $\{l_3, l_{42}, l_1\}$ in P_2, combination $\{l_9, l_{13}, l_{26}\}$ in P_3, combination $\{l_8, l_3, l_9, l_7\}$ in P_{13}, etc. The outages of transmission channels will weaken the power flow transmission from sending-end to receiving-end partitions. For example, the outage of combination $\{l_{28}, l_{38}\}$ disconnects the power sending of partition I and further increases the power sending burden of partition II. It will lead to excessive stress on transmission lines l_3, l_{42} and l_1, and hence induce the subsequent outages.

TABLE II
THE PROPORTIONS OF FCS CONTAINING DIFFERENT NUMBER OF COMMON CFPS

CFP Number	FC Prop.	Average loss(MW)
0	28.887%	0
1	39.250%	1470.67
2	22.792%	2237.79
3	7.873%	2373.58
4	1.198%	1851.38

TABLE III
THE NUMBER OF IDENTIFIED FINAL CFPS UNDER DIFFERENT FC NUMBERS

FC Number	1,000	2,000	5,000	7,500	10,000
CFP Number	23	41	48	49	49
Component Involved	29	36	39	39	39
FC Number	20,000	40,000	60,000	80,000	100,000
CFP Number	50	50	50	50	50
Component Involved	39	39	39	39	39

(iii) Critical CFPs reflect system weakness related to generator, e.g. P_7, P_{12}. Such critical CFPs help to find out which combinations of sequential ongoing line outages of generators would be vulnerable and lead to blackouts. For example, according to P_7, the outage of ongoing line l_{46} will increase the power supply from generator connected to l_5, which will increase the transmission pressure on l_5 and cause l_5 overloaded due to the limited line capacity.

4) Discussion of Common CFP : Common CFP is a CFP mined from a CFP set θ by Algorithm 2, and θ is a set of original CFPs sharing the same load cost and satisfying (7).

Similarity J_{th} in (7) is used to classifies original CFPs into several CFP sets θ, and λ_{th} is used to mine common CFP from θ. J_{th} and λ_{th} are two user-defined parameters. Lower (or larger) value of $J_{th} \in (0, 1]$ will mine fewer (or more) number of final CFPs. The value of $\lambda_{th} \in \{1, 2, \ldots, |\theta|\}$ determines which common components and outage causalities in θ can be reserved. Parameters are set to $J_{th} = 0.45, \lambda_{th} = 2$ in the cases. The proposed method identifies 50 final CFPs that can represent all the 71 original CFPs.

Furthermore, Table II presents proportions of FCs containing different number of common CFPs, where the value of 'FC Prop.' column represents the proportion of FCs containing the corresponding number of common CFPs; and the value of 'Average loss' column represents the average load loss of corresponding FCs. It indicates that the 50 final CFPs can cover 71.113% FCs. 28.887% FCs contain none CFP with accumulative load loss 0MW, which will not aggravate system blackouts. 39.250% FCs contains minimal one CFP, and 1.198% FCs contains maximal 4 CFPs. Though the occurrence frequency of multiple CFPs in one FC is relatively low (i.e. 7.873% FCs contain 3 CFPs and 1.198% FCs contain 4 CFPs), the average load loss of such FCs is normally more severe, which corresponds to the large-scale blackouts and needs more attention. It can be validated that cascading failure process is dominated by CFPs. The occurrence of different CFPs will continuously aggravate system load loss and finally forming different propagation paths.

TABLE IV
COMPUTING PERFORMANCE OF FC GENERATION AND CFP MINING

FC Number	100,000	50,000	50,000
Per FC Generation time /s	0.21		0.28
CFP Number	50	50	67
CFP mining time /min	2.53	1.27	1.76

TABLE V
CRITICAL CFPS WITH HIGH PRI IN IEEE 118-BUS TEST SYSTEM

No.	CFP	PRI	FC Prop.
1	$(l_{30}, l_{108}, l_{116}, l_{119})$	3.62×10^{-2}	51.31%
2	$(l_{30}, l_{104}, l_{107}, l_{108}, l_{116}, l_{119})$	2.60×10^{-2}	23.56%
3	$\langle(l_{30}, l_{32}, l_{96}, l_{102}, l_{104}) \rightarrow l_{105} \rightarrow l_{106}\rangle$	6.82×10^{-3}	17.58%
4	$(l_{128}, l_{148}, l_{151}, l_{152}, l_{153})$	3.67×10^{-3}	3.76%
5	$\langle(l_{32}, l_{37}) \rightarrow l_{36} \rightarrow l_{54} \rightarrow l_{96} \rightarrow l_{30} \rightarrow l_{48} \rightarrow l_{45}\rangle$	1.52×10^{-3}	2.12%
6	$(l_{114}, l_{115}, l_{116}, l_{119}, l_{127})$	1.02×10^{-3}	1.59%
7	$\langle l_{157} \rightarrow l_{128} \rightarrow l_{148} \rightarrow l_{152} \rightarrow l_{153}\rangle$	9.73×10^{-4}	1.03%
8	$(l_{45}, l_{48}, l_{54}, l_{108}, l_{116}, l_{119})$	9.55×10^{-4}	1.24%
9	$\langle(l_{19}, l_{22}, l_{37}) \rightarrow l_{18}\rangle$	9.34×10^{-4}	1.95%
10	$(l_{105}, l_{106}, l_{107}, l_{108}, l_{116}, l_{119})$	8.30×10^{-4}	3.27%

5) Computing performance: Table III presents the number of identified final CFPs under different FC numbers. It shows that CFP number increases with FCs number. When FC number increases to a certain level (i.e. $\geq 20,000$), the identified CFPs are completely same with the same CFP number and involved components.

To further analyze computing performance of the proposed method, we record the FC generation time and CFP mining time under three scenarios:1) Scenario1: 100,000 FCs under normal load level; 2) Scenario2: 50,000 FCs under normal load level; 3) Scenario3: 50,000 FCs under stressed load level (load level increased by 20%). Comparisons are shown in Table IV.

It shows that per FC generation time is longer in stressed system than in normal system (e.g. $0.28s$ in stressed Scenario3 and $0.21s$ in normal Scenario1 and 2), since the stressed system can generate more complex FC containing more outage causalities. Furthermore, CFP mining time in stressed system is longer than in normal system (e.g. $1.76min$ in stressed Scenario3 and $1.27min$ in normal Scenario2), when mining the same number of FCs. It is because more complex FCs contain more CFPs and will consume longer CFP mining time. CFP mining time is longer when mining more number of FCs (e.g. $2.53min$ in Scenario1 with 100,000 FCs and $1.27min$ in Scenario2 with 50,000 FCs), though the identified final CFPs are completely same. It is because more FCs will consume longer time for constructing more search trees by the proposed method.

B. IEEE 118-bus test system

The IEEE 118-bus test system includes 54 generators, 186 components (177 transmission lines and 9 transformers) and 4878 MW of load in total. 100,000 FCs triggered by N-2 random contingencies are used in the proposed method.

1) Critical CFP Identification: The proposed method totally identifies 128 final CFPs, involving 73 components. For better demonstration, define the first 10 final CFPs as the critical CFPs with PRI (by (13)) greater than 2% of the maximum PRI, as shown in Table V, where the value of 'FC Prop.' column

represents the proportion of FCs containing the corresponding critical CFP shown in the 'CFP' column. The No.9 CFP P_9 $\langle(l_{19}, l_{22}, l_{37}) \rightarrow l_{18}\rangle$ means there are 6 CFPs in the actual propagation paths with A_3^3 permutation forms of (l_{19}, l_{22}, l_{37}). Since the topology structure, generation/load distribution and power flow distribution in IEEE 118-bus test system are more complex, the high-risk component combinations and outage causalities become more diverse, consequently causing the diversified forms of CFP in length and in outege sequence.

Table V indicates that some critical CFPs have high PRI because of high occurrence frequency(e.g. P_1, P_2), some critical CFPs have high PRI because of severe load losses (e.g. P_7). Some CFPs share the same subsequence (e.g. P_1 and P_2 share subsequence $(l_{30}, l_{108}, l_{116}, l_{119})$) but belong to different CFPs, because of different loss cost $\eta(P_1) \neq \eta(P_2)$. Such CFPs will form different propagation paths and cause different consequences to blackouts, which corresponds to the overlapped CFP occurrence type introduced in Section II-B.

Furthermore, the first component outage of a specific CFP is more likely to be triggered in the initial stage, and induces the rest component outages of the CFP in subsequent stages, finally forming a propagation path containing the whole CFP. For example, P_9 can be expanded into 6 specific CFPs. The outage stage distribution of P_9 indicates that for all 1.95% proportion of FCs containing P_9, the first component $l_{19}/l_{22}/l_{37}$ is triggered in the initial stage, and the rest components of P_9 are sequentially induced in subsequent stages. For a specific CFP $\langle l_{19} \rightarrow l_{22} \rightarrow l_{37} \rightarrow l_{18}\rangle$ of P_9, l_{19} is tripped as one of initial contingency, and l_{22}, l_{37}, l_{18} are sequentially tripped because of the outage causality contained. In addition, for the critical CFP with high occurrence frequency (i.e. 51.31% proportion of FCs containing P_1), the frequency of the CFP appears in the initial stage is relatively low (3.91% proportion of FCs). It is because such CFP is the common CFP (mined using Algorithm 2) representing a set of CFPs, which is the subsequence of these CFPs. For a specific original CFP P_1 represented, its first component is more likely tripped in the initial stage, and induces the subsequence in subsequent stages. It is a general scenario for the set of CFPs, and we mine the common subsequence $(l_{30}, l_{108}, l_{116}, l_{119})$ from the CFP set as a common CFP.

2) Topological Characteristic: The topological characteristics of the critical CFPs in IEEE 118-bus test system are significant, which can either isolate the network into sub-networks or reveal the important transmission channels. As shown in Fig.6, the identified critical CFPs (e.g. $P_2, P_3, P_4, P_5, P_6, P_7, P_9$ in Table V), which are sequences of component outages, can clearly isolate the network into different sub-networks. For a combination of components that can isolate power system, it undertakes important burden of power flow transmission. The outages of some components in the combination will increase transmission stress of other components in the combination, the excessive stress will more likely to further induce other components in combination to be tripped. In other words, the occurrence frequency of outage causalities between the components of our identified critical CFPs are higher than other components. For example, a specific CFP $\langle l_{19} \rightarrow l_{37} \rightarrow l_{22} \rightarrow l_{18}\rangle$ of P_9, there exist

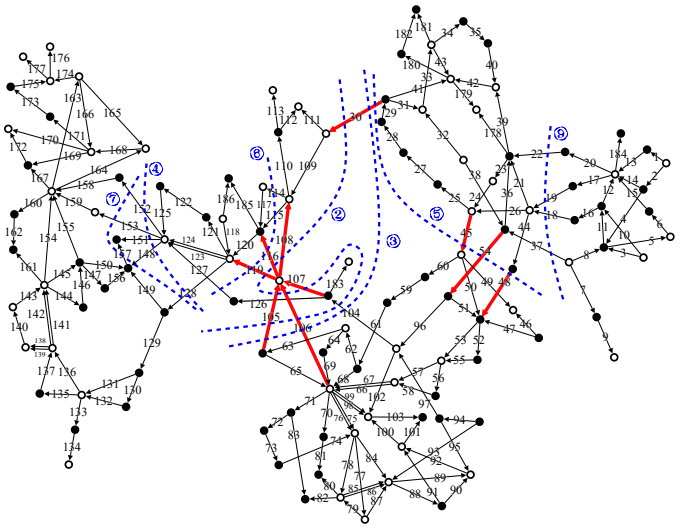

Fig. 6. Critical CFPs identified with 100,000 FCs in IEEE 118-bus test system. The circles represent the nodes in system, where solid circles are load nodes and hollow circles are generator nodes. The link between two nodes represents 186 components numbered as $l_1, \ldots, l_{186}$, and the arrow of link represents initial power flow direction. The components of critical CFPs (e.g. P_1, P_8, P_{10}) that represent important transmission channels, are denoted by red bold lines. The components of critical CFPs (e.g. $P_2, P_3, P_4, P_5, P_6, P_7, P_9$) that can isolate the network, are linked together by blue dashed lines.

outage propagation paths such as $\langle \cdots \rightarrow l_{19} \rightarrow l_{37} \rightarrow l_{22} \rightarrow l_{18} \cdots \rangle (\lambda = 402)$, $\langle \cdots \rightarrow l_{19} \rightarrow l_{37} \rightarrow l_{128} \rightarrow l_{22} \rightarrow l_{18} \cdots \rangle (\lambda = 24)$, $\langle \cdots \rightarrow l_{19} \rightarrow l_{37} \rightarrow l_{119} \rightarrow l_{127} \rightarrow l_{120} \rightarrow l_{107} \rightarrow l_{22} \rightarrow l_{18} \cdots \rangle (\lambda = 6)$. The component gap $g_2^{(P_9)}$ has values 0,1,4. It indicates that though the component outages l_{19} and l_{37} may induce other component outages outside P_9 because of flow redistribution, the outage probabilities of components l_{22} and l_{18} inside P_9 are relatively higher. It is because l_{22} and l_{18} will have higher flow transmission stress especially after l_{19} and l_{37} have been tripped.

In addition, some critical CFPs(e.g. P_1, P_8, P_{10}) reveal important transmission channels of the network. As shown in Fig.6, the outages of such CFPs will weaken the connections between partitions of the system and reduce the main transmission channels for power flows. For example, the outages of components $l_{30}, l_{108}, l_{116}, l_{119}$ in P_1 decrease the transmission of power flow between the sending-end partition (located at the right side of blue dashed line labeled "②" in Fig.6) and the receiving-end partition (located at the left side of line "②"), leading to the subsequent cascading failure propagation with load losses for the equilibrium of power system.

3) CFP-based Mitigation: For verifying the effectiveness of identification results, a CFP-based mitigation is proposed based on critical CFPs in Table V. According to [15], information of CFP can be used for wide area protection schemes, such as blocking relays of components in critical CFPs.

Similar to [15], we assume that 90% of the outages of overloaded components are due to the operation of zone 3 relays. Specifically, reduce the propagation probability of outage causality by 90%. According to outages of current fault state, predict the possible propagation paths according to CFPs, and block the next possible components by reducing the component outage possibilities to 10% temporari-

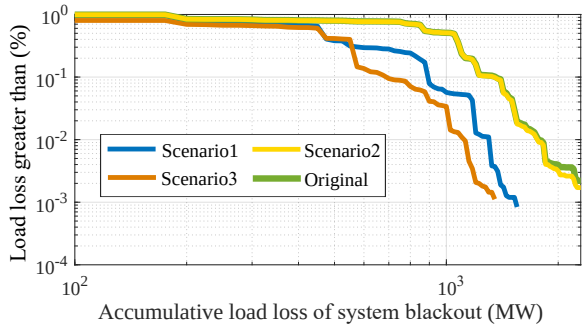

Fig. 7. The CCDs of accumulative load loss of system blackout under CFP-based mitigation measures in IEEE 118-bus test system. The horizontal axis indicates the load losses of blackouts, and the vertical axis indicates the probability of load losses greater than the value of the corresponding horizontal-axis. Scenario1, scenario2 and scenario3 represent mitigation based on P_1, P_7 and 10 critical CFPs in Table V respectively.

ly. For example, for fault state $\langle l_{17}, l_{157}, l_{37}\rangle$, according to P_5, P_7, P_9 in Table V, the outage probability of components $l_{32}, l_{36}, l_{128}, l_{19}, l_{22}, l_{18}$ are set to 10% of the original value in the next stage temporarily.

To demonstrate the effectiveness of CFP-based mitigation, we consider three scenarios with CFP-based mitigation: 1) P_1 in Table V; 2) P_7 in Table V; 3) 10 critical CFPs in Table V, and compare the results with original scenario without mitigation. Complementary cumulative distribution (CCD) is adopted to present the probability distribution of accumulative load losses of system blackouts in each scenario. Fig.7 presents the CCDs for above 4 scenarios. The CCDs under different scenarios are obtained based on 30,000 FCs by cascading failure simulation. Note that blackouts with high load losses may be unreliable due to small simulation samples. For ensuring the reliability of Fig.7, we need to make sure the data on CCDs are all valid. According to [29], with $c = 95\%$ confidence level and $\alpha = 1.5$ accuracy factor, the minimum sample number requirement is 26. Then, the maximum accumulative load loss D_{max} of blackout satisfying sample requirement ($\geq$ 26) is calculated as $1550, 2375, 1325$MW for scenarios1,2,3, respectively. The probabilities of blackout events with $\geq D_{max}$ are grouped together and the average value of these load losses is calculated for representing the blackout events whose accumulative load losses are $\geq D_{max}$ [15].

Fig.7 illustrates that the CFP-based mitigation, which weakens critical CFPs by blocking the outage propagation of components in CFPs, performs well in mitigating. In Scenario1, P_1 exists in 51.31% FCs, the weakening of P_1 will greatly reduce the occurrence frequency of P_1 in most propagation paths and reduce the resulting accumulative load losses. The mitigation based only on this CFP can significantly inhibit the occurrence of severe consequence blackouts. For comparison, scenario2 is the mitigation based on P_7, which has relatively higher loss cost, but the occurrence frequency 1.03% in FCs is lower. Its mitigation alone shows little inhibition on blackouts, whereas its high risk index indicates that the pattern do exist in propagation paths and have specific topological characteristic to arouse our attention. It can function as important power flow transmission lines for preventing the system from further isolating, when other transmission channels are no longer available. Scenario3 is the mitigation considering all 10 critical CFPs, which indicates that the mitigation considering multiple CFPs can reduce the probability of large-scale blackouts more effectively. This suggests that the comprehensive consideration of different critical CFPs can reduce the probability of severe consequence blackouts to a greater extent.

V. Conclusions

In this paper, we proposes the concept of cascading failure pattern (CFP) and a cascading failure pattern identification method based on sequential pattern mining approach for mining CFPs from massive fault chains (FCs). CFP is a combination of sequential component outages with outage causalities, which is vulnerable to be tripped and will aggravate system load loss, playing dominating role in propagation. Critical CFPs are identified for studying how CFPs influence cascading failure propagation, which well represent the significant topological characteristics, such as network isolation and important transmission channels. The proposed method mines CFPs from FCs using cascading failure pattern mining algorithm, mine common CFPs with common components and consequences using common CFP mining algorithm and further identify the critical CFPs. Verifications on IEEE 39-bus and 118-bus test systems illustrate the effectiveness of identification, and the CFP-based mitigation measures are suggested based on critical CFPs for mitigating the risk of cascading failure blackouts.

References

[1] U.S.-Canada Power System Outage Task Force, "Final report on the August 14, 2003 blackout in the United States and Canada: causes and recommendations," System:238, 2004.

[2] European Regulators Group for Electricity and Gas, "Final report the lessons to be learned from the large disturbance in the European power system on the 4th of November 2006," 1C35, 2007.

[3] The Enquiry Committee. Ministry of Power, "Report of the enquiry committee on grid disturbance in northern region on 30th July 2012 and in northern, eastern and north-eastern region on 31st july 2012," Government of India.Tech. Rep, August 2012.

[4] National Grid interim report, "Interim Report into the Low Frequency Demand Disconnection (LFDD) following Generator Trips and Frequency Excursion on 9 Aug 2019," The institute of Engineering and Technology. Tech. Rep., August 2012.

[5] A. Wang, Y. Luo, G. Tu, and P. Liu, "Vulnerability assessment scheme for power system transmission networks based on the fault chain theory," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 442–450, Feb 2011.

[6] H. Guo, C. Zheng, H. H.-C. Iu, and T. Fernando, "A critical review of cascading failure analysis and modeling of power system," *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 9 – 22, 2017.

[7] S. Mei, F. He, X. Zhang, S. Wu, and G. Wang, "An improved opa model and blackout risk assessment," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 814–823, May 2009.

[8] S. Mei, Y. Ni, G. Wang, and S. Wu, "A study of self-organized criticality of power system under cascading failures based on ac-opf with voltage stability margin," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1719–1726, 2008.

[9] M. A. Rios, D. S. Kirschen, D. Jayaweera, D. P. Nedic, and R. N. Allan, "Value of security: modeling time-dependent phenomena and weather conditions," *IEEE Trans. Power Syst.*, vol. 17, no. 3, pp. 543–548, Aug 2002.

[10] J. Song, E. Cotilla-Sanchez, G. Ghanavati, and P. D. H. Hines, "Dynamic modeling of cascading failure in power systems," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2085–2095, May 2016.

[11] I. Dobson and B. Carreras, "A loading-dependent model of probabilistic cascading failure," *Probab. Eng. Inf. Sci.*, vol. 19, 05 2004.

[12] J. Kim and I. Dobson, "Approximating a loading-dependent cascading failure model with a branching process," *IEEE Transactions on Reliability*, vol. 59, no. 4, pp. 691–699, Dec 2010.

[13] X. Wei, J. Zhao, T. Huang, and E. Bompard, "A novel cascading faults graph based transmission network vulnerability assessment method," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2995–3000, May 2018.
[14] Z. Ma, F. Liu, C. Shen, Z. Wang, and S. Mei, "Fast searching strategy for critical cascading paths toward blackouts," *IEEE Access*, vol. 6, pp. 36 874–36 886, 2018.
[15] J. Qi, K. Sun, and S. Mei, "An interaction model for simulation and mitigation of cascading failures," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 804–819, March 2015.
[16] J. Qi, J. Wang, and K. Sun, "Efficient estimation of component interactions for cascading failure analysis by em algorithm," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3153–3161, May 2018.
[17] P.D.Hines, I.Dobson, and P.Rezaei, "Cascading power outages propagate locally in an influence graph that is not the actual grid topology," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 958–967, March 2017.
[18] Z. Zhang, S. Huang, Y. Chen, S. Mei, R. Yao, and K. Sun, "An online search method for representative risky fault chains based on reinforcement learning and knowledge transfer," *IEEE Trans. Power Syst.*, pp. 1–1, 2019.
[19] Y. Liu, S. Huang, S. Mei, and X. Zhang, "A fast searching method for cascading failure pattern based on prefixspan algorithm," in *2018 International Conference on Power System Technology (POWERCON)*, 2018, pp. 345–350.
[20] N. R. Mabroukeh and C. I. Ezeife, "A taxonomy of sequential pattern mining algorithms," *ACM Comput. Surv.*, vol. 43, no. 1, pp. 3:1–3:41, Dec. 2010.
[21] R. Srikant and R. Agrawal, "Mining sequential patterns: Generalizations and performance improvements," *EDBT, Lecture notes in computer science. vol. 1057*, vol. 1057, 03 1996.
[22] Y. Goo, K. Shim, M. Lee, and M. Kim, "Protocol specification extraction based on contiguous sequential pattern algorithm," *IEEE Access*, vol. 7, pp. 36 057–36 074, 2019.
[23] S. Vijayalakshmi, V. Mohan, and S. Raja, "Mining constraint-based multidimensional frequent sequential pattern in web logs," *Eur. J. of Sci. Res.*, vol. 36, 10 2009.
[24] Y.-L. Chen, M.-H. Kuo, S.-y. Wu, and K. Tang, "Discovering recency, frequency, and monetary (RFM) sequential patterns from customers purchasing data," *Electron. Commer. R. A.*, vol. 8, pp. 241–251, 10 2009.
[25] A. Wang, Y. Luo, G. Tu, and P. Liu, "Vulnerability assessment scheme for power system transmission networks based on the fault chain theory," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 442–450, Feb 2011.
[26] B. A. Carreras, D. E. Newman, and I. Dobson, "Determining the vulnerabilities of the power transmission system," in *2012 45th Hawaii International Conference on System Sciences*, 2012, pp. 2044–2053.
[27] R. Yao, S. Huang, K. Sun, F. Liu, X. Zhang, S. Mei, W. Wei, and L. Ding, "Risk assessment of multi-timescale cascading outages based on markovian tree search," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2887–2900, July 2017.
[28] Matpower 7.0, [Online], https://matpower.org/2019/06/20/matpower-7-0-released/.
[29] I. Dobson, B. A. Carreras, and D. E. Newman, "How many occurrences of rare blackout events are needed to estimate event probability?" *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3509–3510, 2013.

Lu Liu (S'19) received the B.E. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2018, where she is currently working toward the M.S. degree in electrical engineering. Her research interests include power system cascading failure and stability analysis.

Hao Wu (M'10) received the Ph.D. degree in electrical engineering from Zhejiang University, Hangzhou, Zhejiang, China. His current research interests include power system operation and stability, cascading failure, uncertainty analysis, load modeling and polynomial chaos expansion.

Linzhi Li (S'17) received the B.E. and Ph.D. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2014 and 2020, respectively. His research interests include cascading failure simulation, risk assessment and management.

Danfeng Shen (S'19) received the B.S. and M.S. degrees in electrical engineering from Tongji University, Shanghai, China, in 2015 and 2018, respectively. He is currently working toward the Ph.D. degree in electrical engineering with Zhejiang University, Hangzhou, China. His research interests include power system uncertainty quantification and polynomial chaos expansion.

Feng Qian received the Ph.D. degree from the China Electric Power Research Institution, Beijing, China. He is currently a Senior Engineer with the Electric Power Dispatching and Control Center of Guangdong Power Grid Company, Guangzhou, China. His research interests include the analysis, control and operation of power systems.

Junlei Liu received the B.S., M.S. and Ph.D. degrees from the South China University of Technology, Guangzhou, China. He is currently an Engineer with the Electric Power Dispatching and Control Center of Guangdong Power Grid Company, Guangzhou. His research interests include the analysis of safety and stability, and the control of reactive power and voltage for power systems.