IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 22, 2018, accepted May 29, 2018, date of publication June 12, 2018, date of current version July 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2846022

Fast Searching Strategy for Critical Cascading Paths Toward Blackouts

ZHIYUAN MA, (Student Member, IEEE), FENG LIU, (Member, IEEE), CHEN SHEN, (Senior Member, IEEE), ZHAOJIAN WANG, (Student Member, IEEE), AND SHENGWEI MEI, (Fellow, IEEE)

China State Key Laboratory of Power System and Generation Equipment, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China

Corresponding author: Feng Liu (lfeng@tsinghua.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51621065 and in part by HQ sponsored S&T project of State Grid Corporation of China: Study on the development scale, bearing capability, and evolutionary principle of balanced structure for power grids.

ABSTRACT The analysis of severe cascading blackouts is an essential issue in power grid planning and operation. As there are a tremendous number of possible cascade paths in a large power system, it is very challenging to determine the critical cascading paths that may have catastrophic consequences. This paper proposes a complex network theory based methodology to determine such critical cascading paths with high efficiency. To this end, we construct an improved interaction graph to cope with the situation where the $(N-1)$ criterion is not satisfied during the follow-up cascade propagation after an initial fault. Subsequently, based on the graph, we derive a modified PageRank model to fast rank the influence of individual transmission lines on blackout risks. Further, leveraging the results of influence ranking, we devise an efficient strategy for searching critical cascading paths. Then, we derive an unbiased probability estimation method for individual cascading paths and the blackout. Simulations carried out on the IEEE 39-bus system, the IEEE 118-bus system, and a real-world 1122-bus system in China show that our method can enhance the searching efficiency by up to three orders of magnitude compared with standard Monte Carlo approaches, demonstrating its potential for cascading blackout analysis in large-scale power systems. Results also verify that the proposed probability estimation method is unbiased, and it can provide an efficient way for probability estimation of the blackouts with very low probability but high losses.

INDEX TERMS Power grid, search strategy, cascading path, blackouts, probability estimation.

I. INTRODUCTION

Blackouts in power systems are recognized as rare events, with low probability but high impact, which can cause huge losses. The probability distribution of blackout size is usually heavy-tailed, indicating that the risk of large rare blackouts cannot be neglected [1]. Accordingly, a method that can efficiently identify such large and rare events is desirable in planning and operation, to facilitate the mitigation of blackout risks. Thus far, conducting large-scale simulations is still the main approach in analyzing cascading blackouts. However, this is practically restrictive due to the "combinational explosion" of possible cascade paths. For example, consider a rare event that occurs with a probability of 10^{-9}. Even if the computer can do 1000 runs per second, one still needs 11 days to capture such a rare event, from the point of view of statistics. In this regard, this paper aims at proposing a method to speed up cascading blackout simulation to determine those paths that could lead to large blackouts in power system operation, facilitating a high-efficiency blackout analysis.

In the literature, studies on cascading blackout simulations in power grids can be roughly divided into two categories: sampling-based approaches [2]–[7] and non-sampling-based approaches [8], [9]. Regarding the former, Monte Carlo (MC) methods are among the most popular. While they are advantageous in modeling complicated systems, a high computational burden largely restricts their practical application [10]. To improve the efficiency, several variance reduction approaches have been proposed, such as the importance sampling and splitting method. The concept behind the importance sampling method is to use a proposal failure probability distribution instead of the original one, such that the probability of rare events can be significantly increased. Empirically, importance sampling, in combination with variance reduction techniques, can achieve a speedup of 2∼4 in [2] and 1∼2 orders of magnitude in [3]. The splitting method follows an alternative idea. It divides blackout simulation into stages, using the number of tripped lines or the percentage of demand loss. In [5], a splitting method is applied to compute the probability of events where more than half the total load is shed. It is further developed by optimizing the parameter settings in the simulation [6]. The optimal choice of splitting stages is also addressed in [4] and [7]. Simulation results of these works indicate that the splitting method can be 100 times faster than standard MC. Regarding non-sampling-based approaches, various contingency/state combination selection techniques have been proposed. In particular, the Random Chemistry (RC) method presented in [8] and [9] demonstrates a high efficiency in identifying $(N-k)$ contingencies that may initiate large blackouts. It is at least an order of magnitude faster than the standard MC method. Nevertheless, room for improvement in efficiency of searching for cascading paths that lead to large blackouts still exists.

Note that, when sampling-based methods are applied, lots of computation resources are wasted on sampling non-influential states, which are unlikely to result in severe blackouts. This indicates that one could capture blackout events in a more efficient manner by focusing on the most influential states/components that exacerbate the propagation of cascading failures. Following this line of thought, this paper proposes a fast searching strategy for cascade paths that would lead to severe blackouts by ranking the influence of transmission lines in a power grid. Lines with high ranking score are regarded as highly influential lines (HILs), which play a core role in boosting cascading propagation of failures, and could be remarkably helpful in searching for critical cascading paths that may lead to blackouts. In this context, the main contributions of this paper are as follows:

1) We improve the interaction graph presented in [11], so that the PageRank algorithm [11] can be applied to rank the influence of lines during the propagation process of cascading failures. Crucially, it enables the influence ranking algorithm to consider the situation where the $(N-1)$ criterion is not satisfied during the follow-up cascade propagation after an initial fault.
2) Based on the branching process model [12], we devise a method to statistically simulate cascade propagation. Guided by the ranking results from 1), the algorithm can efficiently search for critical cascading paths leading to blackouts, up to three orders of magnitude faster than the standard MC method.
3) In terms of the proposed searching strategy, we further derive an unbiased probability estimation method for individual cascading paths and the blackout.

The remainder of this paper is organized as follows. Section II states the general problems associated with and the procedure of simulating cascading blackouts. The improved PageRank-based algorithm for fast ranking the influence of transmission lines is given in Section III. The generation of subsequent cascading failures is discussed in Section IV. Section V exhaustively presents the searching algorithm. The unbiased probability estimation method for blackouts is given in Section VI. Test results are presented in Section VII. Finally, the conclusions are drawn in Section VIII.

II. PROBLEM DESCRIPTION

A. PROCESS OF CASCADING FAILURES

Typically, cascading events comprise initiating and propagating events, which can be grouped into different generations. Initiating events could be short circuits due to contact of a transmission line with trees, operational errors, or vandalism. This is followed by subsequent generations of propagating events, which cause the system state to change and weaken. In combination, these processes are a cascading failure. Mathematically, this can be formulated as a Markovian tree (MT) [13], [14], which is shown in Fig.1. Each node in Fig.1 represents a system state, and the label on each node refers to the set of tripped lines. Accordingly, a cascade path is defined as the sequence of tripped lines in each generation, such as $(Z_1^{(l)}, Z_2^{(ls)}, \cdots)$. Moreover, the nodes marked with grey color in Fig.1 refer to influential events, which are subjected to the corresponding HILs.

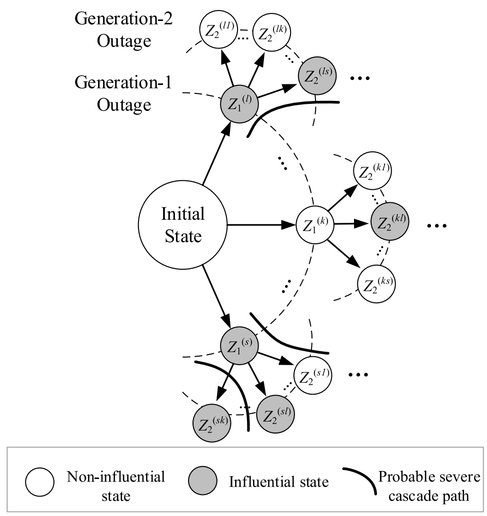


FIGURE 1. General propagation of cascading failure.

B. MOTIVATION

It has been well recognized that a small set of HILs in power systems exist, which may exacerbate cascade propagation, while others contribute much less [15], [16]. For the sake of clearly demonstrating this interesting (also important) pattern, we carried out a cascading blackout simulation on the IEEE 39-bus system using the Monte Carlo method. The statistical results are shown in Fig. 2.

In a total of 460,000 simulation runs, only 1,146 blackouts result in power losses greater than 1,000 MW. Among these

1,146 blackouts, 696 cascade paths (60.73% of the total) are induced by the top 10 HILs, as shown in zone 1 of Fig.2. For the surplus blackouts triggered by the other 36 non-influential lines, there are 281 cascades whose generation-2 outages are triggered by the top 10 HILs, accounting for 24.52% of the total (Fig.2, zone 2). Among the blackouts causing power losses greater than 1,500 MW, the initial or generation-2 outages of 64.91% and 21.83% blackouts, respectively, are triggered by the top 10 HILs. Similar phenomena are observed in the case of blackouts whose power losses are greater than 2,000 MW.

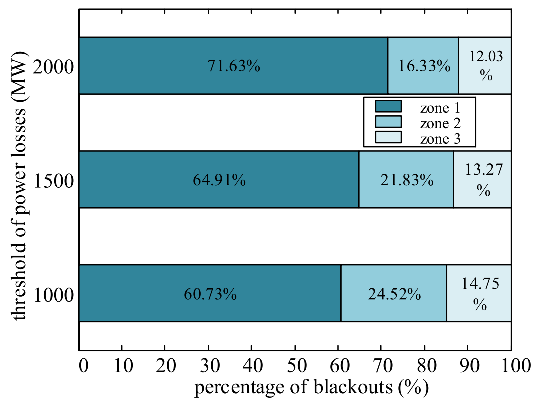


FIGURE 2. **Distribution of blackouts.**

The above example empirically demonstrates that the critical cascading paths are dominated by those HILs, particularly at the initial states. This appealing feature indicates that identifying such HILs may greatly help reduce the search space for determining the critical cascading paths. Following this line of thought, this paper aims at deriving a methodology to accelerate critical cascading path searching by leveraging the results of influence ranking of transmission lines.

C. FRAMEWORK

The general outline of the cascading path searching procedure is summarized in Fig.3, where $N_{\max}$ refers to the total number of simulation runs. In the procedure, three issues are critical: 1) determining the failure probabilities of individual components, 2) determining the failed components at the child generation and 3) recovering the true probability distribution of blackouts. Regarding the first issue, one could assign larger probabilities to the more influential components, to accelerate the critical paths searching. As for the second issue, one needs to determine how the failure propagates in the rest of the grid. With regard to the third issue, one needs to re-scale the simulation results to recover the consequent failure probability distribution triggered by an $(N-1)$ contingency. In this regard, the paper is focused on addressing three critical problems:

P1: How to fast rank the influence of transmission lines on cascading blackout risk with high accuracy?

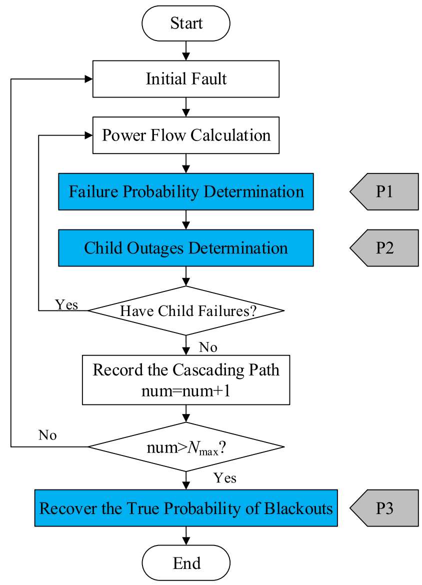


FIGURE 3. **General outline of the cascading path searching procedure.**

P2: How to select child failures to simulate subsequent propagation of a cascading failure?

P3: How to re-scale the simulation results to recover the true probability distribution of cascading blackouts?

Regarding the first problem, in the literature, graph-based approaches have been primarily utilized to evaluate the importance of components in power grids [12], [17]–[20]. Among these graph-based approaches, the eminent PageRank algorithm is well known for its capability of quickly and accurately identifying the most influential nodes in a directed weighted network [21]–[25]. It has been deployed in our prior work [11], in combination with the interaction graph representation of cascading failures, to devise a fast algorithm to screen out the most vulnerable lines in power grids. However, this rudimentary work is built on the assumption that the $(N-1)$ criterion is satisfied; this may not be true during the propagation of cascading failure. In this paper, we extend this work to facilitate influence ranking of lines in successively weakened or degraded system states caused by cascading failure [26]. The key challenge is to extend the interaction graph representation of cascading failures to cope with the situation where the $(N-1)$ criterion is not satisfied. Then the PageRank algorithm can be deployed to rank the influence of each line during the follow-up cascade propagation after an initial fault. This problem will be addressed in Section III.

As for the second problem, several models have been proposed to study cascading failures, such as the CASCADE model [27], the branching process model [28]–[31], the hidden failure model [32], the OPA model, and its variants [33]–[35], to name a few. Among them, the branching process model can describe the statistical features of the average propagation and how the number of outages increases in propagation. It has been further used to estimate the probability distribution of blackout size, such as the load shed [30], [31] or the number of line outages [28]–[31]. It is demonstrated that parameters of the branching process model can adequately depict the average propagation intensity in a statistical manner, with only a small number of cascading failure simulations required [30]. Motivated by this salient feature, this paper deploys the branching process model to statistically determine the number of line failures in each child generation, triggered by parent failures. On the other hand, leveraging the results of influence ranking, the failure probabilities of individual lines are assigned in proportion to their rank scores. This problem will be discussed in detail in Section IV.

As for the third issue, following the idea of sequential importance sampling (SIS) suggested in [3], we devise an unbiased probability estimation method for blackouts in terms of the proposed searching strategy. See Section VI for details.

III. INFLUENCE RANKING BASED ON PAGERANK

In this section, we address the problem **P1**. Specifically, the interaction graph proposed in [11] is improved to enable the PageRank algorithm to rank the influence of lines during the propagation of a cascading failure after an initial fault, where the $(N-1)$ criterion is not satisfied. To this end, we first depict the impact of line outages on hidden failures in terms of actual power fluctuation (see Eqs. (5)∼(6) for details), instead of equivalent power increment, as treated in [11]. Then, we modify the standard PageRank model to cope with the improved interaction graph (IIG).

A. IMPROVED INTERACTION GRAPH

For a power system with n lines, denote the set of lines by $\mathcal{N}$. When the hidden failure is considered, an IIG with $(n+1)$ nodes is derived, shown in Fig.4. Here the filled circles denote lines in the system and the dashed circle represents the hidden failure node. It is assumed that the IIG is directed and weighted. The probability that a line is tripped due to hidden failure is modeled as a monotone function of the line flow [32], which is given in Fig.5. Here, p stands for the probability of mis-operation and $P_{\max}$ is the line limit.

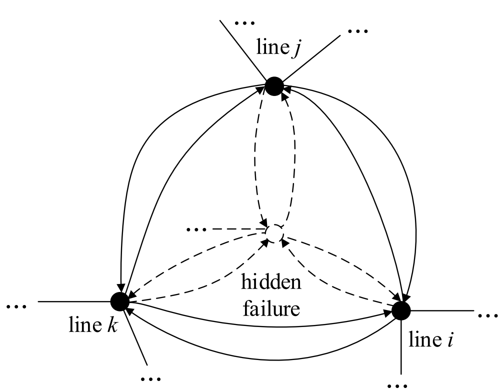


FIGURE 4. Diagram of IIG.

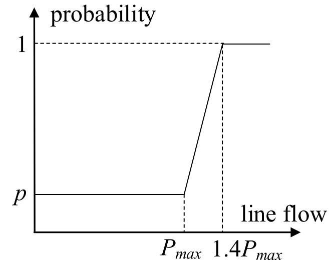


FIGURE 5. Hidden failure model.

The IIG can be described by a $(n+1)\times(n+1)$ matrix $\boldsymbol{X}$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{K} & \boldsymbol{b} \\ \boldsymbol{a}^T & 0 \end{bmatrix} \tag{1}$$

where the dimensions of $\boldsymbol{K}, \boldsymbol{b}, \boldsymbol{a}^T$ are $(n\times n)$, $(n\times 1)$, and $(1\times n)$ respectively. The entries x_{ij} denote the weights of links from node j to i, which are defined as follows.

1) For $1 \le i \le n$, $1 \le j \le n$, we have $x_{ij} = k_{ij}$, where k_{ij} is defined as

$$k_{ij} := \begin{cases} \dfrac{\Delta P_{ij}}{M_i} : & i \neq j \\ 0 : & i = j \end{cases} \tag{2}$$

In (2), ΔP_{ij} denotes the power increase on line i caused by the outage of line j. M_i is the security margin of line i. In this situation, x_{ij} denotes the influence of line j on line i.

2) For $1 \le i \le n$, $j = n+1$, we have $x_{ij} = b_i$, where b_i is defined as

$$b_i := \frac{1.4P_{i\max} - P_i}{M_i} \tag{3}$$

In (3), P_i indicates the power on line i and $P_{i\max}$ is the line limit of line i. Here, $x_{i(n+1)}$ is the influence of hidden failure on line i.

3) For $i = n+1$, $1 \le j \le n$, if the $(N-1)$ criterion is satisfied, we have $x_{ij} = (a^T)_j$. $(a^T)_j$ is defined as

$$(a^T)_j := \sum\nolimits_{i=1, i\neq j}^{n} \frac{\mathbf{Pr}(|P_i^j|)\cdot b_i}{n-1} \tag{4}$$

In (4), P_i^j is the active power on line i after line j is tripped. $\mathbf{Pr}(|P_i^j|)$ stands for the tripping probability of line i caused by the outage of line j, which can be obtained by Fig.5. $x_{(n+1)j}$ denotes the expected influence of line j on hidden failures via other lines.

4) For $i = n+1$, $1 \le j \le n$, if the $(N-1)$ criterion is not satisfied (during the follow-up cascade propagation after an initial fault), we have $x_{ij} = (a^T)_j$. $(a^T)_j$ is

$$(a^T)_j := \sum\nolimits_{i=1, i\neq j}^{n} \frac{\mathbf{Pr}(|P_i^j|)\cdot \tilde{b}_i}{n-1} \tag{5}$$

where

$$\tilde{b}_i := \begin{cases} \frac{|P_i^j|-P_i}{M_i} & |P_i^j| > 1.4P_{i\max} \\ b_i & \text{otherwise} \end{cases} \tag{6}$$

The key difference between this construction and that in [11] is 4), which accounts for the expected effect of line tripping on hidden failures when the system state violates the $(N-1)$ criterion. As shown in (6), if $|P_i^j|$ exceeds 1.4 times its maximum, the actual power fluctuation on i is used to depict the impact from hidden failures.

B. MODIFIED PAGERANK MODEL

Based on interaction matrix $\boldsymbol{X}$, we give the modified PageRank model

$$\left[\boldsymbol{R}^{(k+1)}\right]^T = \left[\boldsymbol{R}^{(k)}\right]^T \boldsymbol{Y} \tag{7}$$

where

$$\boldsymbol{R} := \begin{bmatrix} \widehat{\boldsymbol{R}} \\ r \end{bmatrix} \tag{8}$$

$$\boldsymbol{Y} := \boldsymbol{\Phi}\widetilde{\boldsymbol{X}} \tag{9}$$

$$\widetilde{\boldsymbol{X}} := \begin{bmatrix} \alpha\boldsymbol{K} & \left((1-\alpha)\sum_{i=1}^{n} a_i\right)\boldsymbol{b} \\ \boldsymbol{a}^T & \varepsilon \end{bmatrix} \tag{10}$$

$$\boldsymbol{\Phi} := \begin{bmatrix} \widehat{\boldsymbol{\Phi}} & \boldsymbol{0} \\ \boldsymbol{0}^T & \varphi \end{bmatrix} \tag{11}$$

Here, $\boldsymbol{R}^{(k)}$ is the *influence vector* at iteration k. $\boldsymbol{Y}$ is the *normalized interaction matrix*, i.e., the sum of entries in each row is equal to one. When (7) converges, the equilibrium, denoted by $\boldsymbol{R}^*$, is the final influence vector. The $(n \times 1)$ column vector $\widehat{\boldsymbol{R}}$ is called *line influence vector*, whose i^{th} element (denoted by $\widehat{R}_i$) is the influential degree of line i. r is the influence score of hidden failures. $\widetilde{\boldsymbol{X}}$ is a weighted partitioned matrix. α and $(1-\alpha)$ weight the interactive influence among lines and the influence of hidden failure on lines, respectively. Here, α is set as 0.85 [21]. $\varepsilon > 0$ is a small number, much lower than components of $\boldsymbol{a}$. $\boldsymbol{\Phi}$ is a diagonal matrix with $\Phi_{ii} := 1/\sum_{j=1}^{n+1}\tilde{x}_{ij}$, where $\widehat{\boldsymbol{\Phi}}$ is the n^{th}-order principal minor determinant of $\boldsymbol{\Phi}$. Specifically, we have

$$\varphi = \frac{1}{\varepsilon + \sum_{i=1}^{n} a_i} \tag{12}$$

Similar to [11], it is easy to prove that $\boldsymbol{Y}$ is *stochastic*, *irreducible*, and *aperiodic*. These three properties guarantee $\boldsymbol{R}$ to converge to a unique influence vector $\boldsymbol{R}^*$. The stationary line influence vector is denoted by $\widehat{\boldsymbol{R}}^*$.

C. PAGERANK-BASED RANKING ALGORITHM

The PageRank algorithm can be applied to obtain the stationary influence vector $\boldsymbol{R}^*$, leading to a PageRank-based algorithm to rank the influence of lines. The overall algorithm is given in Algorithm 1.

Algorithm 1 PageRank-Based Influence Ranking

Input: Active power flow data
Output: Influential degree order of lines

1: **for** i=1:n **do**
2: **for** j=1:n **do**
3: **if** $i \neq j$ **then**
4: Obtain ΔP_{ij} by the line outage distribution factor [36].
5: **end if**
6: **end for**
7: **end for**
8: Obtain matrix $\boldsymbol{K}$ by (2).
9: Obtain $\boldsymbol{b}$ by (3) and $\boldsymbol{a}$ by (4)∼(6).
10: Obtain $\boldsymbol{Y}$ by (9)∼(11).
11: Solve (7). Set the starting vector $\boldsymbol{R}^{(0)} = \boldsymbol{1}$ with all entries as 1. Obtain $\boldsymbol{R}^*$ by applying the power method to $\boldsymbol{Y}$.
12: Rank transmission lines by their influence degrees according to $\widehat{\boldsymbol{R}}^*$. A bigger value in $\widehat{\boldsymbol{R}}^*$ indicates a higher influence.

IV. SIMULATING SUBSEQUENT FAILURES

In this section, we focus on resolving problem **P2**, i.e., simulating the subsequent generations of cascading failure by properly selecting child outage lines. Specifically, the branching process model, which is well known for depicting the propagation of cascading failures from a macroscopic point of view [29], is adopted here. It is assumed that each outage at a generation propagates independently to the next generation [12]. While this assumption appears to be idealistic, empirical results match well with many blackout models, such as the OPA model [28] and the CASCADE model [37]. Denote $Z_m^{(d)}$ as the set of line outages at generation m of cascade d. Under this assumption, for each line $i \in Z_m^{(d)}$, the resultant child outages at generation $m+1$ can be determined by the following two steps:

1) Ascertaining the number of child failures, denoted by $\kappa_{m,i}$;
2) Tripping $\kappa_{m,i}$ lines by sampling from the rest of the lines if $\kappa_{m,i} > 0$.

In the following subsections, we exhaustively address the two issues.

A. THE NUMBER OF CHILD FAILURES

1) THE NUMBER OF CHILD FAILURES $\kappa_{m,i}$ TRIGGERED BY LINE i

It is suggested in [12] that the integer number, $\kappa_{m,i}$, follows a Poisson probability distribution, i.e.,

$$\mathbf{Pr}(\kappa_{m,i} = K) = \frac{(\lambda_m^i)^K}{(K)!}e^{-\lambda_m^i}, \quad (K = 0, 1, 2\ldots) \tag{13}$$

where, λ_m^i refers to the average number of line outages triggered by the failure of line i at generation m. Consequently, the number of child outages $\kappa_{m,i}$ resulting from line i can be determined by sampling from the Poisson distribution (13). The details can be found in reference [12].

Obviously, there will be no line tripped for $\kappa_{m,i} = 0$, and one line tripped for $\kappa_{m,i} = 1$. However, if $\kappa_{m,i} > 1$, multiple lines will be tripped simultaneously. We consider the scenarios that multiple lines might be tripped together due to the following reasons. In practice, there does exist such a situation that multiple lines fail together. For example, when a short circuit happens on a bus, all the lines connected to the bus will be tripped together. On the other hand, in cascading failure simulations, it is common to use "generation division" to group the events which happen very closely into the same generation [26], [29], [39], [40].

2) POISSON PARAMETER λ_m^i

Notably, before applying (13) to ascertain the number of child outages $\kappa_{m,i}$, we need to estimate the Poisson parameter λ_m^i for each line i and each generation m. In [12], λ_m^i is statistically given by

$$\lambda_m^i = C_m^i / P_m^i \tag{14}$$

where, P_m^i is the number of times that line i appears as a parent outage at generation m; C_m^i refers to the total number of *effective children*,[1] which results from the failure of line i at the respective generation(s).

Furthermore, λ_m^i is simply divided into two cases [12]: λ_0^i for the initial generation; λ_{1+}^i for all subsequent generations. This means that the number of child failures produced by line i at the subsequent generations follows an identical distribution. In this paper, we continue to adopt this rationale.

3) THE NUMBER OF CASCADES NEEDED TO ESTIMATE λ_0^i AND λ_{1+}^i

For the sake of estimating λ_0^i and λ_{1+}^i, a certain number of cascades, T, need to be performed in advance. These simulated cascades are called *pilot cascades*. In order to improve computational efficiency, it is desirable to minimize the number of pilot cascades, which has not been addressed in the literature, including [12]. Next, we use a simple heuristic method to adaptively determine the minimum number of pilot simulations, $T_{\min}$.

At the beginning, a certain number of pilot cascades, denoted by t_0, are triggered by $(n-1)$ contingencies, to calculate λ_0^i and λ_{1+}^i for each line $i \in \mathcal{N}$. Then, we gradually increase the number of pilot cascades, until the calculated λ_0^i and λ_{1+}^i converge. We denote the increased number of pilot cascades by ΔT. Then, for the j-th run of pilot cascade simulations, the number of pilot simulations is[2]

$$T_j = [t_0 + \Delta T \times (j-1)] \times n, \quad j = 1, 2, \cdots.$$

At the end of each run j, we calculate $\lambda_0^i(j)$ and $\lambda_{1+}^i(j)$ for every line i. Heuristically, if four successive $\lambda_0^i(j)$ (and $\lambda_{1+}^i(j)$) are close enough (say, less than a given threshold τ), the Poisson parameters λ_0^i and λ_{1+}^i of line i are regarded as converged. Then the corresponding number of pilot cascades for this line can be obtained, denoted by $T_{\min}^i$. When the Poisson parameters for all lines are converged, the pilot simulation ends, and $T_{\min}$ is determined by

$$T_{\min} := \max\left\{T_{\min}^1, T_{\min}^2, \cdots, T_{\min}^n\right\} \tag{15}$$

The calculation process is illustrated in Fig. 6.

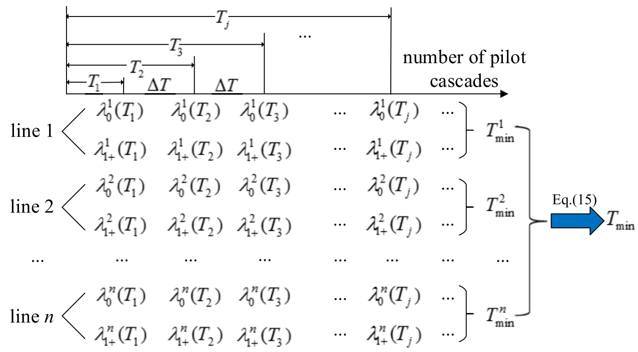


FIGURE 6. Procedure of pilot cascade simulations.

B. SAMPLING FAILED LINES AT CHILD GENERATION

In this subsection, we focus on resolving the second issue: for line $i \in Z_m^{(d)}$, if $\kappa_{m,i} > 0$, determining its child outages at the next generation by sampling $\kappa_{m,i}$ lines from the rest of the power grid .

As stated in Section II, if the propagation of cascades can be guided by HILs, it could significantly reduce the searching space, resulting in a high-efficiency searching strategy of critical cascading paths toward blackouts. A straightforward idea is to assign higher failure probabilities to lines with higher influence rank scores. Specifically, we define the tripping probability of line k as

$$\mathbf{Pr}(\text{line } k \text{ fails at generation } m+1) := \frac{\widehat{R}_{m,k}}{\sum_{k \in \Omega_m} \widehat{R}_{m,k}} \tag{16}$$

where, Ω_m is the set of normal lines at generation m and $\widehat{R}_{m,k}$ is the influential degree of line k at generation m, given by the PageRank algorithm proposed in Section III (Algorithm 1).

The failed lines at the child generation can be sampled subject to the probability function (16). Here, we use Roulette selection [38] to realize the sampling. Note that if the number of child outages satisfies $\kappa_{m,i} > 0$, one needs to repeat the Roulette selection $\kappa_{m,i}$ times, as presented in Procedure I.

C. ALGORITHMS

By utilizing $T_{\min}$ pilot cascades, Algorithm 2 presents the approach for estimating λ_i^0 and λ_i^{1+} for each line i in a power grid. Subsequently, for all lines in $Z_m^{(d)}$, the procedure for selecting their child outages at the next generation is outlined in Algorithm 3. Here, $Z_{m+1}^{(d)}$ denotes the set of child outages at generation $m+1$ of cascade d.

[1] For example, if line i and j appear as parent outages at generation m of a particular cascade, and generation $(m+1)$ of the same cascade includes 3 child outages, the effective children for line i and j could be 3/2=1.5.

[2] The multiplier n is due to the fact that there are n lines, and we need to proceed pilot simulations for each of the n lines.

Procedure I: ROULETTE SELECTION METHOD

Step 1: For the z^{th} line l_z of Ω_m ($z = 1, 2\cdots, |\Omega_m|$), calculate its cumulative probability by

$$q(l_z) \quad := \quad \sum_{j=1}^{z} \mathbf{Pr}(\text{line } l_j \text{ fails}) \qquad (17)$$

Here, $|\Omega_m|$ denotes the cardinality of Ω_m.

Step 2: If $q(l_1) > \gamma$, select line l_1; otherwise select line l_z, s.t. $q(l_{z-1}) < \gamma \leq q(l_z)$. Here, γ is a uniformly distributed random number satisfying $\gamma \in [0, 1]$.

Step 3: Repeat Step 2 $\kappa_{m,i}$ times. All the selected lines compose the child outages triggered by line i.

Algorithm 2 Estimating Poisson Parameters

Input: Active power flow data

Output: λ_0^i and λ_{1+}^i for each line i in the power grid

1: Increase the number of pilot cascades, until the Poisson parameters for all lines are converged.
2: Determine the minimum number of pilot cascades $T_{\min}$ by (15).
3: Obtain λ_0^i and λ_{1+}^i for each line i in the power grid.

Algorithm 3 Simulating Subsequent Failures

Input: Active power flow data
Poisson parameter λ_0^i and λ_{1+}^i for each line $i \in Z_m^{(d)}$

Output: $Z_{m+1}^{(d)}$

1: For each line $i \in Z_m^{(d)}$, determine how many child outages $\kappa_{m,i}$ exist by sampling from (13)~(14).
2: Call **Algorithm 1** to obtain the influence value of the lines.
3: For each line $i \in Z_m^{(d)}$, use the Roulette selection method presented in Procedure I to obtain its child outages if $\kappa_{m,i} > 0$.
4: All the child outages compose $Z_{m+1}^{(d)}$.

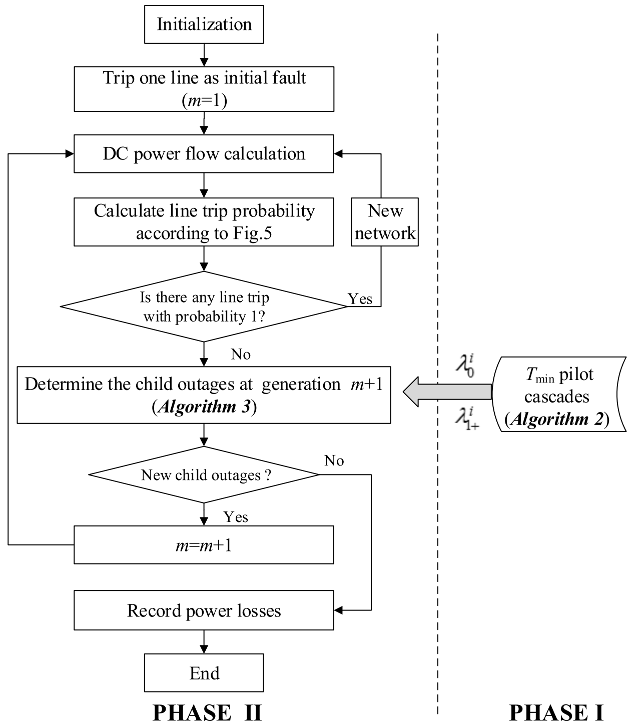


FIGURE 7. Flow chart of fast searching strategy for critical cascading paths.

V. OVERALL SEARCHING ALGORITHM

The overall searching strategy is presented in Fig.7, where Algorithms 2 and 3 are embedded. It involves two phases: PHASE I is processed to estimate λ_0^i and λ_{1+}^i for all lines by calling Algorithm 2; PHASE II is then processed for searching critical cascading paths by leveraging the results of PHASE I. PHASE I is executed only ONCE before running PHASE II. As long as PHASE I is completed, it outputs λ_0^i and λ_{1+}^i of each line i as fixed parameters. When PHASE II is executed, these Poisson parameters can be used directly without any change throughout the follow-up process.

VI. PROBABILITY ESTIMATION OF BLACKOUTS

In this section, we address the problem **P3**. Trip line i as an initial fault, and the probability of a given event A causing the power losses greater than PL_{thr} is defined as $\mu(A, i)$. If the MC method is adopted, $\mu(A, i)$ is estimated by (18) after N_s simulation runs, denoted by $\tilde{\mu}_{MC}(A, i)$.

$$\tilde{\mu}_{MC}(A, i) = \frac{1}{N_s} \sum_{j=1}^{N_s} \delta_{\{h(z_{MC}^j) \geq PL_{\text{thr}}\}} \qquad (18)$$

where, z_{MC}^j is the j^{th} sample of cascading paths; $h(z_{MC}^j)$ stands for the total power loss of sample z_{MC}^j; $\delta_{\{\cdot\}}$ is the indicator function of set $\{\cdot\}$. Here, $\delta_{\{\cdot\}} = 1$ if the condition in set is satisfied; otherwise $\delta_{\{\cdot\}} = 0$.

Equation (18) provides an unbiased estimation on the true probability $\mu(A, i)$. However, it cannot be directly applied to the proposed searching strategy, since PHASE II uses the surrogate failure probabilities to replace the actual ones at each generation of the subsequent cascading failure simulations. Therefore, one has to re-scale the simulation results to recover the actual probability distribution of the blackouts caused by an $(N-1)$ contingency.

Denote the j^{th} samples resulting from PHASE I and PHASE II by z_{ss1}^j and z_{ss2}^j, respectively. Denote the ratio of actual and surrogate probability of these two samples by $\omega(z_{ss1}^j)$ and $\omega(z_{ss2}^j)$, respectively. After obtaining N_s samples by the proposed searching strategy, $\mu(A, i)$ can be estimated by

$$\tilde{\mu}_{SS}(A, i) = \frac{1}{N_s}\Bigg(\sum_{j=1}^{T_{\min}/n} \omega(z_{ss1}^j)\cdot\delta_{\{h(z_{ss1}^j)\geq PL_{\text{thr}}\}} + \sum_{j=(T_{\min}/n)+1}^{N_s} \omega(z_{ss2}^j)\cdot\delta_{\{h(z_{ss2}^j)\geq PL_{\text{thr}}\}}\Bigg) \qquad (19)$$

Specifically, $\omega(z_{ss1}^j)$ in (19) is equal to 1, since PHASE I directly uses the MC method to simulate cascading paths.

As for $\omega(z^j_{ss2})$, it is computed by

$$\omega(z^j_{ss2}) = \frac{p^j_c}{q^j_c} = \frac{\prod_{m=1}^{n^j} \bar{p}^j_{m,m+1}}{\prod_{m=1}^{n^j} \bar{q}^j_{m,m+1}} \tag{20}$$

where

$$\bar{p}^j_{m,m+1} = \prod_{k\in\bar{\Omega}^j_m} \hat{p}^j_{m,k} \prod_{k\in\Omega^j_m} (1-\hat{p}^j_{m,k}) \tag{21}$$

$$\bar{q}^j_{m,m+1} = \prod_{k\in\bar{\Omega}^j_m} \hat{q}^j_{m,k} \tag{22}$$

In (20), p^j_c and q^j_c refer to the actual probability and surrogate probability of a given sample z^j_{ss2}, respectively. n^j is the length of this sample. In (21), $\bar{p}^j_{m,m+1}$ is the actual transition probability from generation m to $m+1$, of the sample z^j_{ss2}. $\bar{\Omega}^j_m$ denotes the set of tripped lines at generation m of the sample z^j_{ss2}. Accordingly, other normal lines compose the set Ω^j_m, and they will be still in operation normally at the $(m+1)$-th generation. $\hat{p}^j_{m,k}$ is the actual tripping probability of line k at generation m of the sample z^j_{ss2}, which can be obtained by employing the hidden failure model in Fig.5. In (22), $\bar{q}^j_{m,m+1}$ describes the surrogate transition probability from generation m to $m+1$, of the sample z^j_{ss2}. $\hat{q}^j_{m,k}$ refers to the surrogate tripping probability of line k at generation m, which can be calculated by (16).

Remark 1: Essentially, the simulation process in PHASE II is the same as a sequence of sequential importance sampling (SIS) steps [3]. Note that, in [3], it has been proved that the SIS method gives an unbiased estimation on the probability distribution of cascading blackouts. Therefore, $\tilde{\mu}_{SS}(A,i)$ in our method (computed by (19)-(22)) is also an unbiased estimator for $\mu(A,i)$.

VII. CASE STUDIES

In this section, we used the standard IEEE 39-bus system and IEEE-118 system to verify the efficacy and efficiency of the proposed methodology. Furthermore, a real-world 1,122-bus power grid in China is used for demonstrate the practicality of our method. All the three systems are set as initially $(N-1)$ secure.

A. VERIFICATION TECHNIQUE

In the case studies, we used the standard MC method as a comparison to demonstrate the efficiency of the proposed searching strategy. Specifically, we employed the improved blackout simulation model [11] incorporating the hidden failure mechanism [32]. The key improvement is that our model allows relay misoperations with certain probabilities on all the operating lines rather than only the lines exposed to incorrect tripping. Therefore, our model is more realistic. The flow chart of MC method is shown in Fig.8.

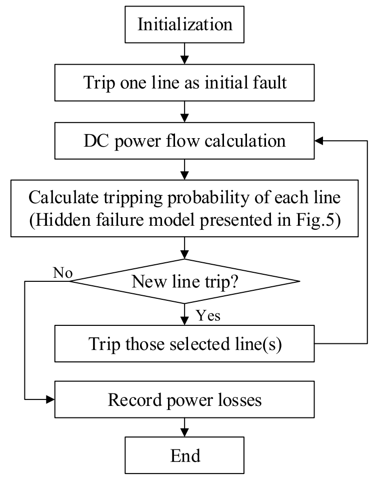


FIGURE 8. Flow chart of the Monte Carlo (MC) method.

Trip line i as the initial fault. The number of cascading paths causing power losses greater than PL_{thr} obtained by the proposed searching strategy and the MC method, are defined as $\Psi_{SS}(i)$ and $\Psi_{MC}(i)$, respectively. To quantitatively measure the improvement in searching efficiency, we introduced an indicator, which is given by

$$SF(i) := \frac{\Psi_{SS}(i)}{\Psi_{MC}(i)} \tag{23}$$

Clearly, a larger SF implies a more significant improvement in searching efficiency.

B. CASE 1: IEEE 39-BUS SYSTEM

In this case, the IEEE 39-bus system was used for testing. This system has 46 transmission lines, 10 generators, and 6,254 MW of load.

1) PILOT CASCADE SIMULATIONS

The datasets of pilot cascades were produced using the simulator presented in Section VII-A. We first determined the minimum number of pilot cascades $T_{\min}$ by using the method proposed in Section IV-A, and then computed λ^i_0 and λ^i_{1+} for each line i in the power grid.

Next, the initial number of simulations t_0 was set to 100. We initiated from $T_1 = t_0 \times n = 100 \times 46 = 4,600$ and set ΔT as 4,600. In addition, the convergence threshold τ was set to 0.1. For each line $i \in \mathcal{N}$, we calculated its Poisson parameters λ^i_0 and λ^i_{1+}. After 23 runs, all the parameters converged (the variance of four successive parameters is less than τ). The minimum number of pilot cascades for each line is shown in Fig. 9. Clearly, the highest value of this histogram is $105,800$, implying that pilot cascades $T_{\min} = 105,800$. Line 10 is taken as an example to illustrate the process of convergence, as shown in Fig.10. Fig.11 shows the final estimation of Poisson parameters λ^i_0 and λ^i_{1+} for some lines.

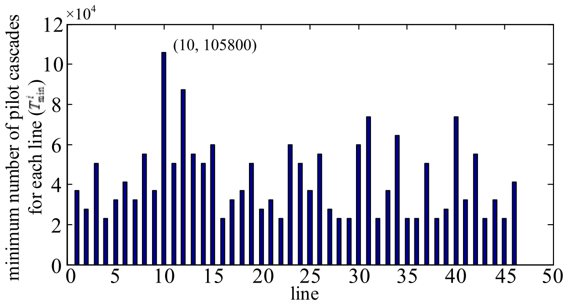


FIGURE 9. Minimum number of pilot cascades for each line.

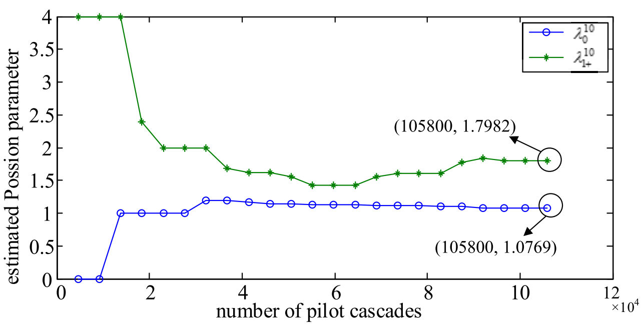


FIGURE 10. Process of parameter estimation (line 10).

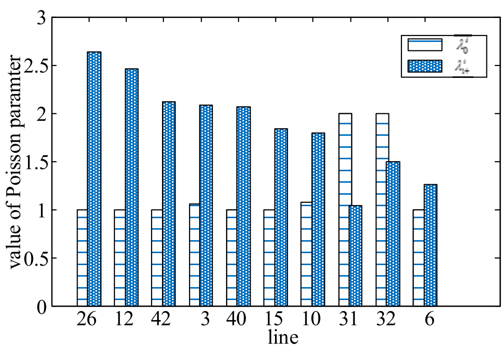


FIGURE 11. Estimation results of Poisson parameters.

2) EFFICIENCY

Next, we compared the efficiency of the proposed searching strategy in simulating severe blackouts with that of the MC method.

For the MC method, we performed 10, 000 cascading simulations for each $(N-1)$ contingency (460,000 simulations in total). For the proposed searching strategy, we first allocated 2,300 pilot cascades for each $(N-1)$ event (105,800 total simulations) to estimate the Poisson parameters of all the lines (PHASE I of searching strategy). Then, the remaining 7, 700 cascading simulations were conducted for each initiating event (354,200 total simulations) to search for critical cascading paths (PHASE II of searching strategy). In this way, the total number of samples in the two methods remains the same. We considered the following three scenarios:

1) **S1:** $PL_{\text{thr}} = 1,000\text{MW}$;
2) **S2:** $PL_{\text{thr}} = 1,500\text{MW}$;
3) **S3:** $PL_{\text{thr}} = 2,000\text{MW}$.

For example, consider initial faults at line 5 and line 46, respectively. The growth of identification results obtained using the two aforementioned methods are compared in Fig.12. The results show that the number of blackouts captured using the proposed searching strategy is almost zero when the given computation budget is used to simulate pilot cascades, and then increases linearly with the increase in simulation runs. The growth rate of our proposed method is much faster than that of the MC method. Interestingly, the results also indicate that a vast major of blackouts obtained through the proposed searching strategy are unique, as shown in Tab.1. All these results confirm the high efficiency of our algorithm in searching critical cascading paths toward blackouts.

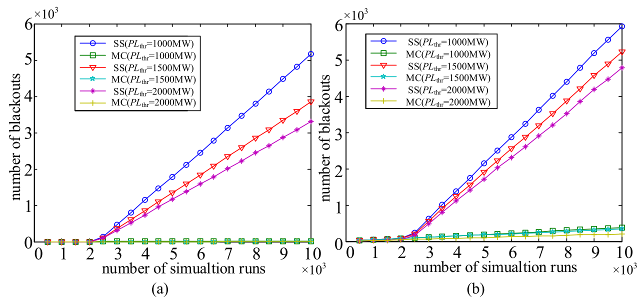


FIGURE 12. Comparison results for initial fault at (a) line 5 and (b) at line 46.

TABLE 1. Percentage of unique blackouts for initial fault at lines 5 and 46.

Initial fault	PL_{thr}	Number of unique blackouts	Number of blackouts	Percentage(%)
line 5	1000(MW)	4753	5162	92.08
	1500(MW)	3794	3849	98.57
	2000(MW)	3273	3307	98.97
line 46	1000(MW)	5577	5839	95.51
	1500(MW)	5017	5127	97.85
	2000(MW)	4640	4733	98.04

The speedup factors for different initial-fault lines are compared in Fig.13. Note that the vertical axis is in logarithm. The horizontal axis refers to the initial fault selected randomly. As shown in Fig.13, most of the results provide a speedup factor of 2$\sim$3 orders of magnitude. Specially, when $PL_{\text{thr}} = 2,000$ MW, the speedup factor for line 34 is infinite since the MC method fails to capture any blackouts (therefore the result is not provided here). In addition, the results indicate that the higher the power loss threshold PL_{thr}, the greater is the speedup factor. It further certificates that the searching strategy has a considerable advantage in effectively identifying severe cascade events.

3) PROBABILITY DISTRIBUTION ESTIMATION OF BLACKOUTS

Trip one line as the initial fault, and estimate the probability distribution of blackouts. For example, consider the initial faults at line 5 and line 13, respectively. As shown in Fig.14, the MC method and the proposed method output almost the same estimations of the complementary

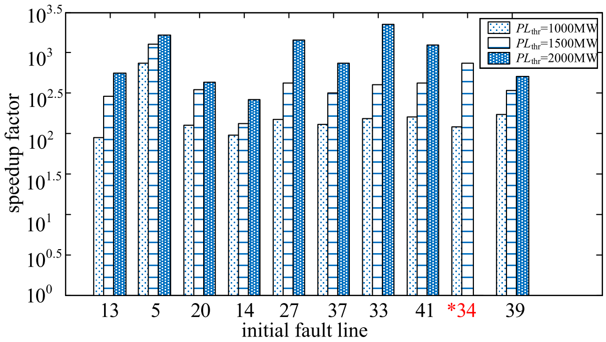


FIGURE 13. Speedup factor for different initial-fault lines.

cumulative probability distribution when the power losses are no more than 2,800MW (Fig.14(a)) and 2,600MW (Fig.14(b)), respectively. It justifies that the proposed searching strategy can achieve an unbiased estimation result.

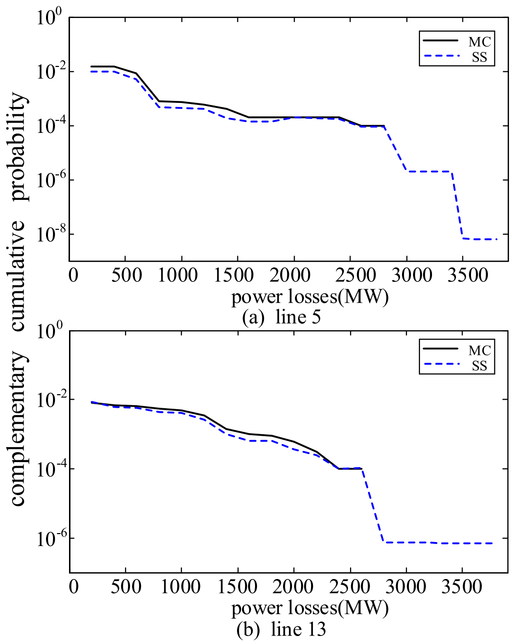


FIGURE 14. Complementary cumulative probability distribution of power losses with the initial fault at (a) line 5 and (b) line 13.

On the other hand, when the power losses are greater than 2,800 MW and the initial fault is set as line 5, the MC method fails to find out any cascading path in 10,000 simulation runs and cannot give the probability distribution estimation of blackouts. In contrast, our searching strategy can find a great number of such critical cascading paths. It is even able to give the probability distribution estimation of the blackouts with a power loss as high as 3,800MW. We further check the probability of every critical path that causes power losses more than 2,800 MW. It is found that the probabilities of individual critical paths range from 1.0227×10^{-59} to 1.2556×10^{-6}. Similar results are obtained when the initial fault is set at line 13.

These results demonstrate that the proposed method provides an efficient way for searching critical paths and estimating the probability distribution of blackouts.

C. CASE 2: IEEE 118-BUS SYSTEM

In this case, the IEEE-118 bus system is used for test. It has 186 transmission lines, 54 generators, and 6,363 MW of load in total.

1) PILOT CASCADE SIMULATIONS

First, the method proposed in Section IV-A was applied to determine the minimum number of pilot cascades $T_{\min}$. The initial value of simulation runs t_0 was then set to 100. Start from $T_1 = t_0 \times n = 18,600$ and let $\Delta T = 18,600$. The acceptable mismatch τ was set to 0.1. The results show that the minimum number of pilot cascades for each $n-1$ contingency used to estimate Poisson parameters is 1,800. Consequently, $T_{\min} = 334,800$. Then, by utilizing those pilot cascades, the estimated Poisson parameters λ_0^i and λ_{1+}^i for each line can be obtained. Owing to limited space, Fig.15 presents only partial results.

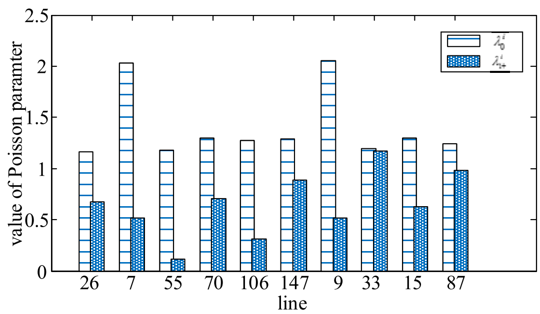


FIGURE 15. Estimation of Poisson parameters.

2) EFFICIENCY

For the MC method, we simulated 10,000 samples of cascades for each initial tripped line. The total number of simulation runs is 1,860,000. For each $(N-1)$ contingency, the computation budgets for PHASE I and II of the proposed searching strategy were set to 1,800 and 8,200, respectively. Similar to Case 1, consider the following three scenarios:

1) **S1:** $PL_{\text{thr}} = 1,000$ MW;
2) **S2:** $PL_{\text{thr}} = 1,500$ MW;
3) **S3:** $PL_{\text{thr}} = 2,000$ MW.

The number of blackouts identified through the two methods are compared in Fig.16. The horizontal axis refers to the initial fault selected randomly. These results indicate that the critical cascading paths searched using the proposed method far outnumber those obtained using the MC, as MC fails to identify any severe blackouts. The results confirm the high efficiency of the proposed searching strategy in capturing cascading blackouts.

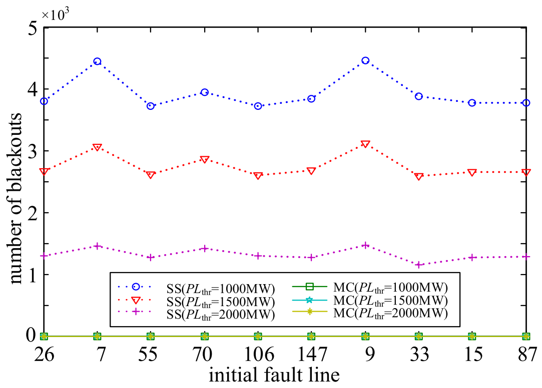


FIGURE 16. Searching results under different initial faults.

3) PROBABILITY DISTRIBUTION ESTIMATION OF BLACKOUTS

We consider the initial faults at line 26 and line 70 as examples. As shown in Fig.17, the estimation results of the complementary cumulative probability distribution given by the proposed searching strategy are almost the same as that given by the MC method, when the power losses are small ($PL_{\text{thr}} <$ 700MW in Fig.17(a) and $PL_{\text{thr}} <$ 900MW in Fig.17(b)). However, the MC method fails to find any critical paths in terms of the initial fault at line 26 when the power losses are larger than 700MW. In contrast, the proposed searching strategy is still effective for capturing such critical paths. By further checking the probability of every critical paths causing power losses more than 700MW, we find that the probabilities of these individual critical cascading paths are extremely small. Specifically, they vary in a broad range from 1.5939×10^{-218} to 2.9398×10^{-8}. The results demonstrate that our estimation method is unbiased and further exhibit the efficacy of the proposed strategy in searching critical cascading paths.

D. CASE 3: 1122-BUS REAL SYSTEM IN CHINA

For test the scalability of the proposed searching strategy, we compare it with the MC method in a large real power grid in China, which has 1,792 transmission lines, 1,122 buses and 51.78GW of load in total.

1) PILOT CASCADE SIMULATIONS

In this case, the initial value of simulation runs t_0 was chosen as 100. Start from $T_1 = t_0 \times n = 179,200$ and let $\Delta T = 179,200$. The acceptable mismatch τ was still set to 0.1. Results indicate that the minimum number of pilot cascades for each $(N-1)$ contingency used to estimate Poisson parameters is 1,100. That is, $T_{\min} = 1,971,200$. Consequently, the estimated Poisson parameters for some lines are shown in Fig.18.

2) EFFICIENCY

For each initial fault, the computation budgets for PHASE I and PHASE II of the proposed searching strategy are

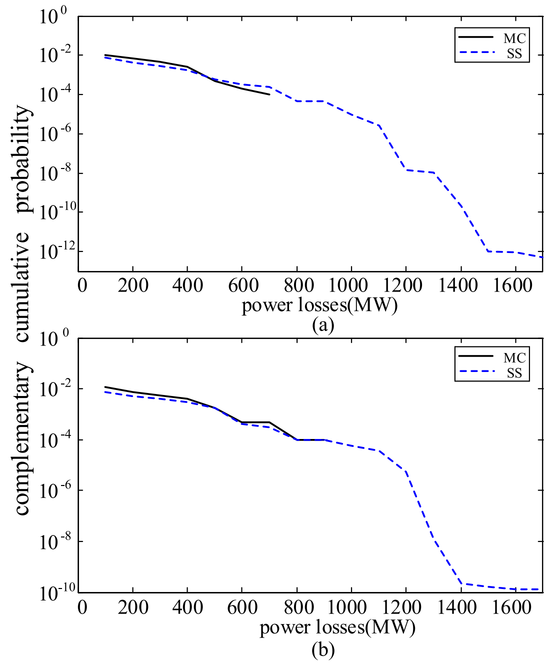


FIGURE 17. Complementary cumulative probability distribution of power losses with the initial fault at (a) line 26 and (b) line 70.

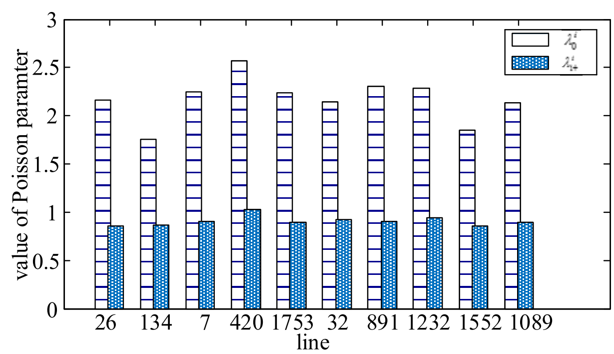


FIGURE 18. Estimation of Poisson parameters.

1,100 and 8,900, respectively. For the Monte Carlo method, we simulate 10,000 samples for each $(N-1)$ contingency. Similar to Case 1 and Case 2, we consider the following three scenarios:

1) **S1:** $PL_{\text{thr}} = 1,000$ MW;
2) **S2:** $PL_{\text{thr}} = 1,500$ MW;
3) **S3:** $PL_{\text{thr}} = 2,000$ MW.

Fig.19 shows the speedup factors for different initial-fault lines, ranging from 10^2 to 10^3. Specifically, when $PL_{\text{thr}} = 2,000$ MW, the speedup factor for line 7 is infinite since the MC method fails to capture any blackouts (therefore the result is not provided here). Moreover, the result indicates that the speedup factor is larger when the power loss threshold PL_{thr} is higher. It verifies that the proposed method indeed has high efficiency in searching critical cascading paths toward blackouts, particularly in large-scale power grids.

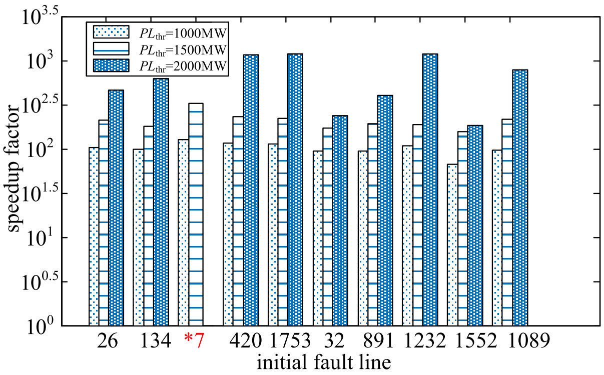


FIGURE 19. Speedup factor for different initial-fault lines.

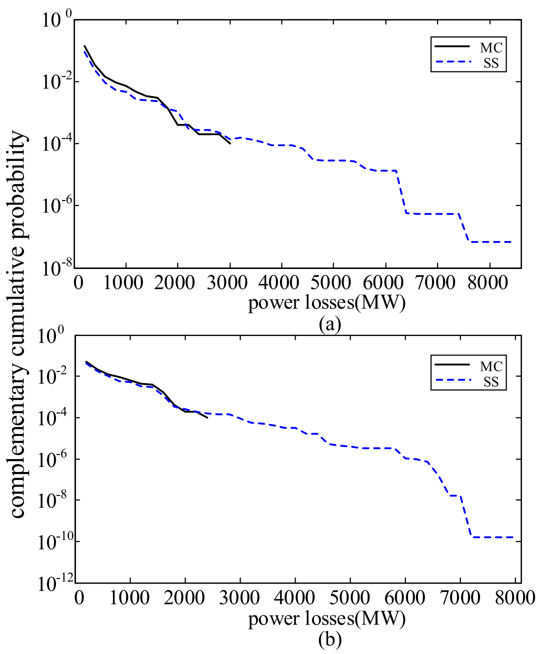


FIGURE 20. Complementary cumulative probability distribution of power losses with the initial fault at (a) line 134 and (b) line 1753.

3) PROBABILITY DISTRIBUTION ESTIMATION OF BLACKOUTS

In this case, we consider the initial faults at line 134 and line 1753 as examples. According to the estimation results shown in Fig.20, the complementary cumulative probability distributions given by these two methods match well when the power losses are small ($PL_{\text{thr}} < 3,000$MW in Fig.20(a) and $PL_{\text{thr}} < 2,400$MW in Fig.20(b)). Moreover, the proposed searching strategy successfully finds out many cascading paths with high power losses even larger than 8,000 MW. It further suggests that the proposed searching strategy can considerably improve the capability of capturing severe blackouts, particularly when the failure probabilities are very low. The result also empirically demonstrates that the proposed method is scalable and practical, which is promising to real applications in large-scale power systems.

VIII. CONCLUSION

In this paper, we have proposed an improved interaction graph to better represent cascading outages of a power grid, particularly when the $(N-1)$ criterion is not satisfied during the follow-up cascading propagation after an initial fault. We then have proposed a PageRank-based algorithm to rapidly rank the influence of lines during the propagation of cascading failures. Further, we have devised a method to simulate subsequent generations of cascading outage based on an improved branching process model. Afterwards, we have devised an unbiased probability estimation method for individual cascading paths and the blackout. The simulation result indicates that our algorithm can significantly accelerate the searching of critical cascading paths toward blackouts in large-scale power systems. It also verifies that the proposed probability estimation method is unbiased and can provide an effective way for probability analysis in cascading blackout simulations with very low probabilities but high losses.

REFERENCES

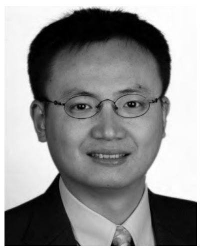
[1] B. A. Carreras, D. E. Newman, and I. Dobson, "North American blackout time series statistics and implications for blackout risk," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4406–4414, Nov. 2016.
[2] Q. Chen and L. Mili, "Composite power system vulnerability evaluation to cascading failures using importance sampling and antithetic variates," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2321–2330, Aug. 2013.
[3] J. Guo, F. Liu, J. Wang, J. Lin, and S. Mei, "Toward efficient cascading outage simulation and probability analysis in power systems," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2370–2382, May 2018.
[4] J. Kim, J. A. Bucklew, and I. Dobson, "Splitting method for speedy simulation of cascading blackouts," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3010–3017, Aug. 2013.
[5] S.-P. Wang, A. Chen, C.-W. Liu, C.-H. Chen, and J. Shortle, "Rare-event splitting simulation for analysis of power system blackouts," in *Proc. IEEE Power Energy Soc. General Meeting*, Jul. 2011, pp. 1–7.
[6] S. P. Wang, A. Chen, C. W. Liu, C. H. Chen, J. Shortle, and J. Y. Wu, "Efficient splitting simulation for blackout analysis," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1775–1783, Jul. 2015.
[7] J. F. Shortle, "Efficient simulation of blackout probabilities using splitting," *Int. J. Elect. Power Energy Syst.*, vol. 44, no. 1, pp. 743–751, Jan. 2013.
[8] M. J. Eppstein and P. D. H. Hines, "A 'random chemistry' algorithm for identifying collections of multiple contingencies that initiate cascading failure," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1698–1705, Aug. 2012.
[9] P. Rezaei, P. D. H. Hines, and M. Eppstein, "Estimating cascading failure risk: Comparing Monte Carlo sampling and random chemistry," in *Proc. IEEE PES General Meeting | Conf. Expo.*, Jul. 2014, pp. 1–5.
[10] I. Dobson, B. A. Carreras, and D. E. Newman, "How many occurrences of rare blackout events are needed to estimate event probability?" *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3509–3510, Aug. 2013.
[11] Z. Ma, C. Shen, F. Liu, and S. Mei, "Fast screening of vulnerable transmission lines in power grids: A PageRank-based approach," *IEEE Trans. Smart Grid*, to be published.
[12] P. D. H. 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, Mar. 2016.
[13] Z. Wang, A. Scaglione, and R. J. Thomas, "A Markov-transition model for cascading failures in power grids," in *Proc. 45th Hawaii Int. Conf. Syst. Sci. (HICSS)*, Jan. 2012, pp. 2115–2124.
[14] R. Yao *et al.*, "Risk assessment of multi-timescale cascading outages based on Markovian tree search," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2887–2900, Jul. 2017.
[15] Y. Yang, T. Nishikawa, and A. E. Motter, "Small vulnerable sets determine large network cascades in power grids," *Science*, vol. 358, no. 6365, p. eaan3184, 2017.

[16] Z. Ma, F. Liu, C. Shen, X. Zhang, S. Mei, and F. Gao, "Speeding up simulations of cascading blackout in power systems by identifying high influential lines," in *Proc. Power Energy Soc. General Meeting*, Jul. 2017, pp. 1–5.
[17] Z. Xiangyu *et al.*, "Identification of key transmission lines in power grid using modified K-core decomposition," in *Proc. 3rd Int. Conf. Electr. Power Energy Convers. Syst. (EPECS)*, Oct. 2013, pp. 1–6.
[18] L. Xu, X.-L. Wang, and X.-F. Wang, "Electric betweenness and its application in vulnerable line identification in power system," in *Proc. CSEE*, vol. 1, 2010, pp. 33–39.
[19] P. D. H. Hines, I. Dobson, E. Cotilla-Sanchez, and M. Eppstein, "'Dual graph' and 'random chemistry' methods for cascading failure analysis," in *Proc. 46th Hawaii Int. Conf. Syst. Sci. (HICSS)*, Jan. 2013, pp. 2141–2150.
[20] 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, Mar. 2015.
[21] A. Langville and C. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton, NJ, USA: Princeton Univ. Press, 2011.
[22] J. Maestre, H. Ishii, and E. Algaba, "Node aggregation for enhancing PageRank," *IEEE Access*, vol. 5, pp. 19799–19811, 2017.
[23] C. Li *et al.*, "CRSPR: PageRank for Android apps," *IEEE Access*, vol. 5, pp. 18004–18015, 2017.
[24] M. Pirouz and J. Zhan, "Toward efficient hub-less real time personalized PageRank," *IEEE Access*, vol. 5, pp. 26364–26375, 2017.
[25] N. Ma, J. Guan, and Y. Zhao, "Bringing PageRank to the citation analysis," *Inf. Process. Manage.*, vol. 44, no. 2, pp. 800–810, Mar. 2008.
[26] I. Dobson and D. E. Newman, "Cascading blackout overall structure and some implications for sampling and mitigation," *Int. J. Elect. Power Energy Syst.*, vol. 86, pp. 29–32, Mar. 2017.
[27] I. Dobson, B. A. Carreras, and D. E. Newman, "A loading-dependent model of probabilistic cascading failure," *Probab. Eng. Inf. Sci.*, vol. 19, no. 1, pp. 15–32, 2005.
[28] I. Dobson, J. Kim, and K. R. Wierzbicki, "Testing branching process estimators of cascading failure with data from a simulation of transmission line outages," *Risk Anal.*, vol. 30, no. 4, pp. 650–662, Apr. 2010.
[29] I. Dobson, "Estimating the propagation and extent of cascading line outages from utility data with a branching process," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2146–2155, Nov. 2012.
[30] J. Qi, I. Dobson, and S. Mei, "Towards estimating the statistics of simulated cascades of outages with branching processes," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3410–3419, Aug. 2013.
[31] J. Qi, W. Ju, and K. Sun, "Estimating the propagation of interdependent cascading outages with multi-type branching processes," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1212–1223, Mar. 2017.
[32] J. Chen, J. S. Thorp, and I. Dobson, "Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model," *Int. J. Elect. Power Energy Syst.*, vol. 27, no. 4, pp. 318–326, May 2005.
[33] B. A. Carreras, D. E. Newman, I. Dobson, and N. S. Degala, "Validating OPA with WECC data," in *Proc. 46th Hawaii Int. Conf. Syst. Sci. (HICSS)*, Jan. 2013, pp. 2197–2204.
[34] 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.
[35] J. Qi, S. Mei, and F. Liu, "Blackout model considering slow process," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3274–3282, Aug. 2013.
[36] J. Guo, Y. Fu, Z. Li, and M. Shahidehpour, "Direct calculation of line outage distribution factors," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1633–1634, Aug. 2009.
[37] J. Kim and I. Dobson, "Approximating a loading-dependent cascading failure model with a branching process," *IEEE Trans. Rel.*, vol. 59, no. 4, pp. 691–699, Dec. 2010.
[38] A. Lipowski and D. Lipowska, "Roulette-wheel selection via stochastic acceptance," *Phys. A, Statist. Mech. Appl.*, vol. 391, no. 6, pp. 2193–2196, 2012.
[39] I. Dobson, B. A. Carreras, and D. E. Newman, "Branching process models for the exponentially increasing portions of cascading failure blackouts," in *Proc. 38th Annu. Hawaii Int. Conf. Syst. Sci. (HICSS)*, Jan. 2005, p. 64a.
[40] H. Ren and I. Dobson, "Using transmission line outage data to estimate cascading failure propagation in an electric power system," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 55, no. 9, pp. 927–931, Sep. 2008.

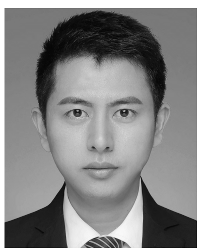
ZHIYUAN MA (S'15) received the B.S. degree from the Shandong University of Science and Technology, Shangdong, China, in 2011, and the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2014, where she is currently pursuing the Ph.D. degree. Her research interests include blackouts and cascading failure analysis.

FENG LIU (M'12) received the B.S. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1999 and 2004, respectively.

From 2015 to 2016, he was a Visiting Associate with the California Institute of Technology, CA, USA. He is currently an Associate Professor with Tsinghua University. He has authored or co-authored over 100 peer-reviewed technical papers and two books, and holds over 20 issued/pending patents. His research interests include power system stability analysis, optimal control and robust dispatch, and game theory and learning theory and their applications to smart grids. He was a Guest Editor of the IEEE TRANSACTIONS ON ENERGY CONVERSION.



CHEN SHEN (M'98–SM'07) received the B.E. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1993 and 1998, respectively. From 1998 to 2001, he held a post-doctoral position with the Department of Electrical Engineering and Computer Science, University of Missouri Rolla, MO, USA. From 2001 to 2002, he was a Senior Application Developer with ISO New England Inc., MA, USA. He has been a Professor with the Department of Electrical Engineering, Tsinghua University, since 2009. He is currently the Director of the Research Center of Cloud Simulation and Intelligent Decision-Making, Energy Internet Research Institute, Tsinghua University. He has authored or co-authored over 150 technical papers and one book, and holds 21 issued patents. His research interests include power system analysis and control, renewable energy generation, and smart grids.



ZHAOJIAN WANG (S'15) received the B.S. degree from the Qiushi Honors College, Tianjin University, Tianjin, China, in 2013. From 2016 to 2017, he was a joint Ph.D. student with the California Institute of Technology, CA, USA. He is currently pursuing the Ph.D. degree in electrical engineering with Tsinghua University, Beijing, China. His research interests include power system distributed optimal control, complex network, and microgrid planning.

SHENGWEI MEI (SM'05–F'15) received the B.S. degree in mathematics from Xinjiang University, Ürümuqi, China, in 1984, the M.S. degree in operations research from Tsinghua University, Beijing, China, in 1989, and the Ph.D. degree in automatic control from the Chinese Academy of Sciences, Beijing, in 1996.

He is currently a Professor with the Department of Electrical Engineering, Tsinghua University. His research interests include power system analysis and control, robust control, and complex systems.