

An Interaction Model for Simulation and Mitigation of Cascading Failures

Junjian Qi, *Member, IEEE*, Kai Sun, *Senior Member, IEEE*, and Shengwei Mei, *Senior Member, IEEE*

Abstract—In this paper the interactions between component failures are quantified and the interaction matrix and interaction network are obtained. The quantified interactions can capture the general propagation patterns of the cascades from utilities or simulation, thus helping to better understand how cascading failures propagate and to identify key links and key components that are crucial for cascading failure propagation. By utilizing these interactions a high-level probabilistic model called interaction model is proposed to study the influence of interactions on cascading failure risk and to support online decision-making. It is much more time efficient to first quantify the interactions between component failures with fewer original cascades from a more detailed cascading failure model and then perform the interaction model simulation than it is to directly simulate a large number of cascades with a more detailed model. Interaction-based mitigation measures are suggested to mitigate cascading failure risk by weakening key links, which can be achieved in real systems by wide area protection such as blocking of some specific protective relays. The proposed interaction quantifying method and interaction model are validated with line outage data generated by the AC OPA cascading simulations on the IEEE 118-bus system.

Index Terms—Blackout, cascading failure, interaction, mitigation, network, power transmission reliability, relay, simulation, wide area protection and control.

I. INTRODUCTION

CASCADING blackouts are complicated sequences of dependent outages which could bring about tremendous economic and social losses [1]–[4]. Large rare cascading blackouts have substantial risk and pose great challenges in simulation, analysis, and mitigation. It is important to study the mechanisms of cascading failures so that the risk of large-scale blackouts may be better quantified and mitigated.

In order to study cascading failures, several models have been proposed, such as CASCADE model [5], branching process model [6]–[8], hidden failure model [9], [10], OPA¹ model [11]–[14], improved OPA model [15], AC OPA model [16], [17], OPA with slow process [18], Manchester model [19], [20], stochastic model [21], dynamic PRA model [22], and influence model [23].

J. Qi and K. Sun are with Dept. of Electrical Engineering & Computer Science, University of Tennessee, Knoxville, TN 37996 (e-mails: junjian.qi.2012@ieee.org and kaisun@utk.edu). S. Mei is with the State Key Laboratory of Power Systems, Dept. of Electrical Engineering, Tsinghua University, Beijing China 100084 (e-mail: meishengwei@mail.tsinghua.edu.cn). This material is based upon work supported partially by the University of Tennessee in Knoxville, the CURENT Engineering Research Center, and NSFC grant-51321005. ©IEEE 2014

¹OPA stands for Oak Ridge National Laboratory, Power Systems Engineering Research Center at the University of Wisconsin, University of Alaska to indicate the institutions collaborating to devise the simulation.

The OPA model [11]–[14] simulates the patterns of cascading blackouts of a power system under the complex dynamics of a growing demand and the engineering responses to failure. Initial line outages are generated randomly by assuming that each line can fail independently. Whenever a line fails, the generation and load is redispatched with DC load flow model and linear programming. If any lines were overloaded during the optimization, then these lines are those that are likely to have experienced high stress, and each of these lines fails independently. The process of redispatch and testing for line outages is iterated until there are no more outages.

The branching process model [6]–[8] can statistically describe the statistical or simulated cascades and provide higher-level statistical information about cascading failures by tracking the numbers of lines outaged and the amounts of load shed. But it does not retain information about the network topology or load flow and also does not attempt to specify how cascades propagate in the system in detail, such as which, where, or why lines outage. The most recent study on the line interaction graph [24] initiates a novel analysis method for cascading failures by considering the interactions of transmission lines and tries to understand cascading failures with models amenable to analysis while keeping the basic physics of power systems.

From the perspective of complex systems the system-level failures are not caused by any specific event but by the property that the components in the system are tightly coupled and interdependent [25]. Thus explicitly studying the interactions between components can help understand the mechanisms of cascading failures, identify the key factors for their propagation, and further propose effective mitigation measures.

In this paper we quantify the interactions between component failures by following the line graph approach in [24]. These interactions can capture the general patterns of the propagation of cascading failures in a system and help better understand why and how cascading failures occur and propagate. Key links between component failures which play important roles in the propagation of cascading failures can also be identified and further be used for determining wide area protection schemes [26]–[29], such as relay blocking under the condition of some specific line tripping, which can secure time to perform remedial controls by a defense system during cascaded events [30]. A cascading failure model called interaction model is further proposed based on these interactions to speed up simulation and to study how component interactions influence cascading failure risks. It is much more time efficient to first quantify the interactions between the

component failures with fewer original cascades from more detailed cascading failure model, such as AC OPA [16], [17], and then perform the interaction model simulation than it is to directly simulate a large number of cascades with a more detailed model.

Besides, topological properties such as small-world [31] and scale-free [32] behaviors have been found in complex networks. But it can be misleading to evaluate the vulnerability of power systems only with topological metrics [33]. In this paper we discuss the property of a directed weighted interaction network generated with simulated cascades from a more detailed cascading failure model which considers the physics of the system such as power flow and re-dispatching rather than directly exploring the property of the network from the topology of the physical system.

The rest of this paper is organized as follows. Section II explains how the interactions between component failures can be quantified. Section III discusses the identification of the key links and key components based on the obtained interactions. Section IV proposes an interaction model by using the quantified interactions and also discusses methods for validating it. Section V proposes methods to determine how many cascades should be simulated and how many cascades should be utilized to quantify the interaction between component failures. Section VI discusses the mitigation measures by weakening key links and the potential application in real systems as a wide area protection scheme. Section VII tests the proposed interaction quantifying method and the interaction model with line outage data generated by AC OPA simulations on IEEE 118-bus system. Finally the conclusion is drawn in Section VIII.

II. QUANTIFYING THE INTERACTIONS BETWEEN COMPONENT FAILURES

In this section the cascades that record cascading failure sequences are used to quantify the interactions between component failures.

For power systems the transmission lines or transformers can be chosen as components and the cascades can come from either statistical utility line outage data or simulations from more detailed cascading failure models. The statistical data can be grouped into different cascades and then into different generations within each cascade based on outages' timing [8]. The simulation data can be generated from OPA model or its variants [11]–[18], which naturally produce line outages in generations or stages; each iteration of the “main loop” of the simulation produces another generation [6], [7]. AC OPA model [16], [17] will be employed in the case studies of this paper to produce cascades for qualifying component interactions. But the data can also be generated from other cascading simulation tools as long as they can be grouped into cascades and generations.

The cascades used for quantifying the interactions between component failures are called original cascades in order to distinguish the simulated cascades from the proposed model in this paper. M original cascades can be arranged as

	generation 0	generation 1	generation 2	...
cascade 1	$F_0^{(1)}$	$F_1^{(1)}$	$F_2^{(1)}$...
cascade 2	$F_0^{(2)}$	$F_1^{(2)}$	$F_2^{(2)}$...
⋮	⋮	⋮	⋮	⋮
cascade M	$F_0^{(M)}$	$F_1^{(M)}$	$F_2^{(M)}$...

where $F_g^{(m)}$ is the set of failed components produced in generation g of cascade m . Each cascade eventually terminates with a finite number of generations when the number of failed components in a generation becomes zero. The shortest cascades stop in generation one by having no outages in generation one and higher generations, but some of the cascades will continue for several or occasionally many generations before terminating.

After obtaining M original cascades we can quantify the interactions between component failures based on all or part of them. Assume $M_u \leq M$ original cascades are utilized to quantify the interactions. Since at first we do not have enough information to determine which components in two consecutive generations have interactions, we assume that there are interactions between any failed component in last generation and that in this generation to guarantee that no interactions will be ignored. Thus for a system with n components, a matrix $\mathbf{A} \in \mathbb{Z}^{n \times n}$ can be constructed, whose entry a_{ij} is the number of times that component i fails in one generation before the failure of component j among all original cascades. Since \mathbf{A} is obtained by using all M_u cascades it does not depend on the order that the cascades are processed.

The assumption based on which \mathbf{A} is obtained actually exaggerates the interactions between component failures since it is not convincing to assert one component interacts with another one only because it fails in its last generation. Therefore, for each failed component in generation one and the following generations the failed component that most probably causes it should be determined.

Specifically, for any two consecutive generations k and $k+1$ of any cascade m , the failure of component j in generation $k+1$ is considered to be caused by a set of failed components in generation k , which can be described as

$$\{i_c | i_c \in F_k^{(m)} \text{ and } a_{i_c j} = \max_{i \in F_k^{(m)}} a_{ij}\}. \quad (1)$$

Note that it is possible that two or more components in generation k are considered as the cause of the failure of component j . When M_u is not large enough this will be more possible because in this case no component has much greater a_{ij} than the others. In the extreme case for which all a_{ij} for $i \in F_k^{(m)}$ are all the same it will become impossible to determine which component more possibly causes the failure of component j and thus all components will be considered as the cause. Then the overestimation of the interaction by \mathbf{A} cannot be well corrected, which will lead to the overestimation of the propagation of cascading failures. This will be discussed further in section VII-A.

An illustration is shown in Fig. 1, in which we show two

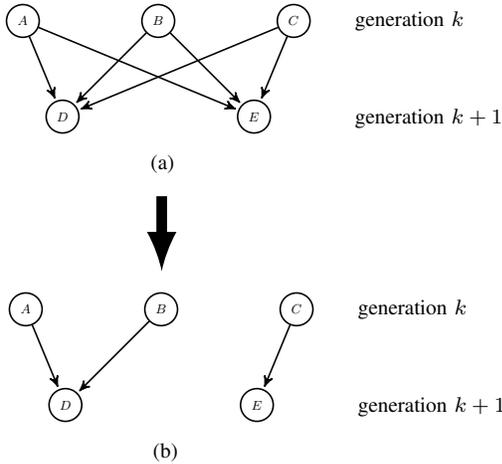


Fig. 1. Illustration for determining the cause of component failures.

consecutive generations of a cascade. If we assume that

$$a_{AD} = a_{BD} = \max_{i \in \{A, B, C\}} a_{iD} \quad (2)$$

$$a_{CE} = \max_{i \in \{A, B, C\}} a_{iE} \quad (3)$$

we can determine the cause of component D failure as component A and B and the cause of component E failure as component C .

What should be emphasized here is that A is not updated when determining the most possible causes for failed components in generation one and the following generations. Therefore, the determination of the component that causes a failed component does not depend on the order that the original cascades are processed but is completely determined by the A matrix.

After determining the cause of any component failure in generation 1 and the following generations for all cascades A can be corrected to be $A' \in \mathbb{Z}^{n \times n}$, whose entry a'_{ij} is the number of times that the failure of component i causes the failure of component j .

The interaction matrix $B \in \mathbb{R}^{n \times n}$ can be calculated from A' . Its entry b_{ij} is the empirical probability that the failure of component i causes the failure of component j , which can be given by

$$b_{ij} = \frac{a'_{ij}}{N_i} \quad (4)$$

where N_i is the number of failures of component i .

The B matrix determines how components interact with each other. The nonzero elements of B are called links. Link $l : i \rightarrow j$ corresponds to B 's nonzero element b_{ij} and starts from component i and ends with component j . By putting all links together a directed network $\mathcal{G}(\mathcal{C}, \mathcal{L})$ called interaction network can be obtained. Its vertices \mathcal{C} are components and each directed link $l \in \mathcal{L}$ represents that a failure of the source vertex component causes the failure of the destination vertex component with probability greater than 0.

III. IDENTIFYING KEY LINKS AND KEY COMPONENTS

The links can vary significantly with respect to their roles in the propagation of cascading failures. In order to distinguish

them and to further identify key links, an index I_l is defined for each link $l : i \rightarrow j$ to be the expected value of the number of failures that are propagated through link l . Note that the failures propagated through link l can be directly triggered by the failure of component i or can be triggered by the failures of components other than i which finally is able to cause component i to fail.

Therefore, in order to calculate I_l the number of failures of its source vertex i , which is denoted by N_i^s , should be set to be N_i , which is the total number of failures among all the original cascades. N_i contains not only the failures in generation 0, which serve as trigger of cascading failures, but also the failures caused by other component failures.

Similar to section II, $M_u \leq M$ original cascades are utilized to quantify the interactions and further to calculate the index I_l . After obtaining the interaction network we can get a directed acyclic subgraph $\mathcal{G}_j(\mathcal{C}_j, \mathcal{L}_j)$ starting with component j from the interaction network \mathcal{G} . The vertices represent the events of component failures and the edges represent causal relations between events. All edges in the subgraph point in the same direction from parent to child due to causality affecting the future and all components are reached exactly once.

We would like to emphasize that for each link there is a unique directed acyclic subgraph which can be extracted from the whole interaction network and is comprised of all the components influenced by this link.

In Fig. 2 we illustrate how the directed acyclic subgraph \mathcal{G}_j (Fig. 2c) can be obtained from the original subgraph (Fig. 2a). Note that i is not in Fig. 2a even if there is a link from j or any other vertex because I_l is defined to indicate the failures link $l : i \rightarrow j$ can cause on the condition that i fails. From Fig. 2a to Fig. 2b we remove the vertices for which there is no path from vertex j to them (H, I , and J denoted by green circles in Fig. 2a) since we would like to quantify the consequences brought about by j and the removed vertices cannot be influenced by j . The links corresponding to the removed vertices (denoted by red arrows in Fig. 2a) are also eliminated.

In Fig. 2b vertex j is at level 0; the vertices that j points to and are not i are at level 1; the vertices that the level 1 vertices point to and are not i or any other vertex in the lower levels are at level 2. Because of the causal relationship between the vertices in two consecutive levels the edges from vertices at a higher level to those at a lower level ($D \rightarrow A$ and $F \rightarrow j$ denoted by red arrows in Fig. 2b) are removed. Also the edges between vertices at the same level ($C \rightarrow B$ denoted by dashed red arrow in Fig. 2b) are neglected since these vertices are considered independent and all fail on the condition of the failure of some component at the last level. Finally we can get a directed acyclic subgraph \mathcal{G}_j (Fig. 2c) for which there is no loop and for each vertex (component) $c \in \mathcal{C}_j$, $c \neq j$ there is exactly one vertex c_s pointing to it.

The expected value of the number of failures of component j given N_i times of component i failure is

$$E_j = N_i^s b_{ij}. \quad (5)$$

For any other component $c \in \mathcal{C}_j$, $c \neq j$, the expected value of the number of failures given the times of its source vertex

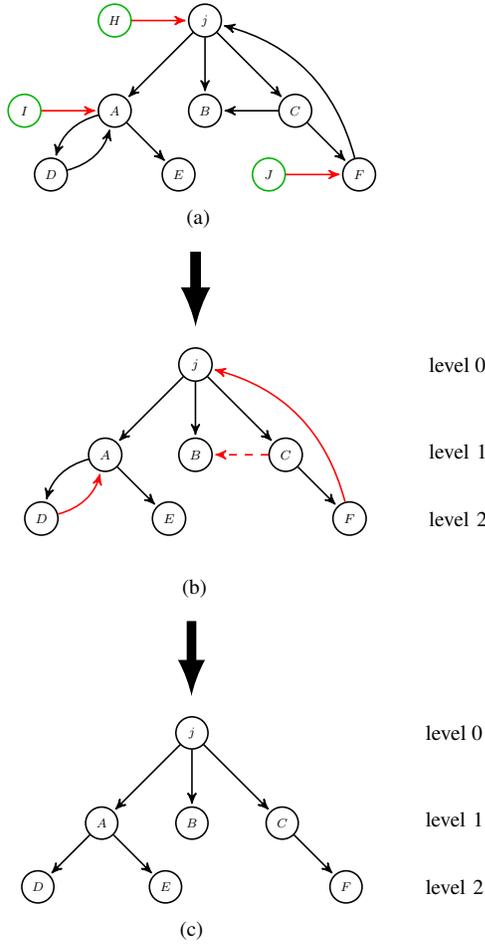


Fig. 2. Illustration for obtaining the directed acyclic subgraph starting with j .

failure is

$$E_c = E_{c_s} b_{c_s j}. \quad (6)$$

All the expected number of component failures in graph \mathcal{G}_j are summated to be I_l as

$$I_l = \sum_{c \in \mathcal{C}_j} E_c. \quad (7)$$

I_l can indicate the contribution of a link to the propagation of cascading failures. The greater the index is, the more important the link is for cascading failure propagation. Thus the links with large I_l can be defined as key links. Specifically, the set of key links \mathcal{L}^{key} are those links whose weights are greater than or equal to a specified fraction of the largest link weight I_l^{max} , that is

$$\mathcal{L}^{\text{key}} = \{l | I_l \geq \epsilon_l I_l^{\text{max}}\} \quad (8)$$

where ϵ_l is taken as a value that is not too close to zero to guarantee that the weights of key links are not much less than the largest link weight.

By taking I_l as weights of the links, we can make the interaction network $\mathcal{G}(\mathcal{C}, \mathcal{L})$ in section II to be a directed weighted network. The vertex out-strength and in-strength of

the interaction network can be defined as follows.

$$s_i^{\text{out}} = \sum_{l \in \mathcal{L}^{\text{out}}(i)} I_l \quad (9)$$

$$s_i^{\text{in}} = \sum_{l \in \mathcal{L}^{\text{in}}(i)} I_l \quad (10)$$

where $\mathcal{L}^{\text{out}}(i)$ and $\mathcal{L}^{\text{in}}(i)$ are respectively the sets of links starting from and ending with vertex i .

The out-strength and in-strength can indicate how much a component influences and is influenced by another one. The components with large out-strength can cause great consequences and thus are crucial for the propagation of cascading failures. Therefore, in a similar way to the key link definition, the set of key components \mathcal{C}^{key} is defined as

$$\mathcal{C}^{\text{key}} = \{i | s_i^{\text{out}} \geq \epsilon_s s_i^{\text{out, max}}\} \quad (11)$$

where $s_i^{\text{out, max}}$ is the largest vertex out-strength among all vertices and ϵ_s is used to guarantee that the out-strengths of the key components are not much less than the maximum out-strength.

IV. INTERACTION MODEL

In this section a cascading failure model called interaction model is proposed by using the tripping probability of each component in generation 0 and the interactions between component failures, which can be obtained either from statistical utility line outage data or simulations generated from OPA model or its variants. After introducing the interaction model we also discuss how it can be validated by comparing its simulated cascades and the original cascades.

As in section II, we assume there are a total of M original cascades available. Note that we do not necessarily need to use all the M cascades but only M_u of them to generate the tripping probability of each component in generation 0 and the interaction matrix since a smaller number of cascades can capture how frequently the components fail in generation 0 and how the component failures interact with each other, especially when any one of the original cascades can be considered as an independent and identical realization of a underlying process.

A. Model Design

It is assumed that all components are initially unfailed and each component fails with a small probability. The component failures in the same generation cause other component failures independently. The flow chart of the proposed model is shown in Fig. 3, in which m_{max} is the number of cascades to be simulated.

The model contains two loops and in each outer loop a cascade is simulated. Specifically, the model is implemented in the following 3 steps.

Step 1) Accidental faults of components

In the k th outer iteration, each component i randomly fail with probability τ_i to simulate accidental faults and the failed components form generation 0 (initial outages) of the simulated cascade. The probability that a component i fails as initial outages can

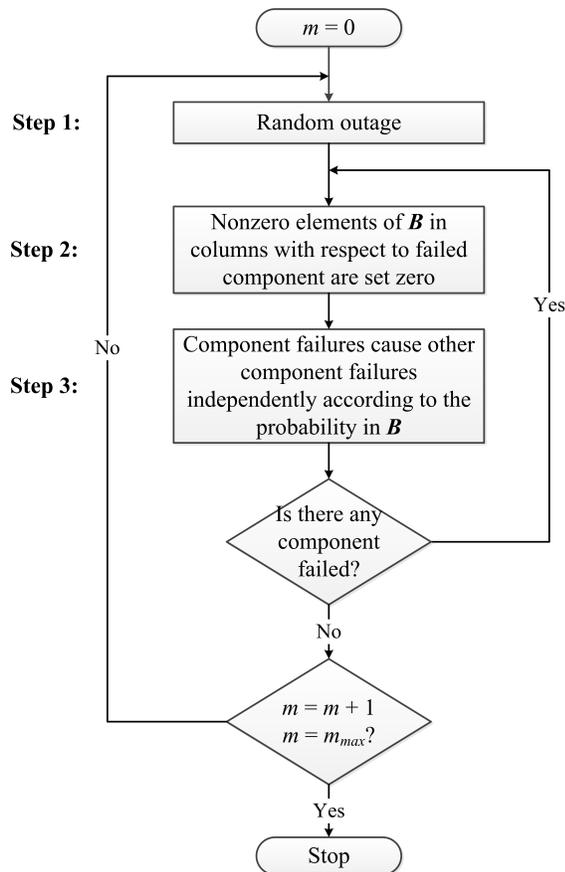


Fig. 3. Flow chart of the interaction model.

be estimated by using the generation 0 component failures of the M_u original cascades as

$$\tau_i = \frac{f_0^i}{M_u} \quad (12)$$

where f_0^i is the number of cascades for which component i fails in generation 0.

Step 2) Corresponding columns of \mathbf{B} are set zero

The columns of \mathbf{B} corresponding to the component failures are set zero since in our model once a component fails it will remain that way until the end of the simulation.

Step 3) Failed components cause other component failures

The component failures in one generation independently generate other component failures. Specifically, if component i fails in this generation it will cause the failure of any other component j with probability b_{ij} . Once it causes the failures of some components, these newly caused component failures will comprise the next generation; then go back to step 2. If no component failure is caused, the inner loop stops.

By using the interaction model described above we can simulate as many cascades as possible (greater than M_u). Although the simulated cascades are generated by utilizing the information of the initial outages and the interactions

contained in the original cascades, they can reveal new rare events due to the high-level probabilistic property of the interaction model, thus helping recover the missing information due to using fewer original cascades. Therefore, as long as the cascades from the interaction model are well validated it can be much more time efficient to first quantify the interactions between the component failures with fewer original cascades from a more detailed cascading failure model and then perform the interaction model simulation than it is to directly simulate a large number of cascades with a more detailed model.

We also would like to emphasize that the proposed interaction model can not only be used for offline study of cascading failures but can also be used for online decision-making support. The interaction matrix can be obtained offline from statistical utility data or simulations of more detailed cascading failure models. It contains important information about the interactions between component failures. By utilizing this information the interaction model has the potential to predict the consequences of events. If something unusual happens in the system the operators can apply the interaction model to quickly find out which components or which areas of the system will most probably be affected so that a fast response can be performed to pull the system back to normal conditions and to avoid or at least reduce the economic and social losses.

B. Validating the Model

In order to validate the proposed interaction model, the simulated cascades from the interaction model are carefully compared with the original cascades by using the following four methods.

- 1) The probability distribution of total line outages of the original and simulated cascades are compared.
- 2) The probability distribution of total line outages of the original and simulated cascades can be estimated with the branching process and the average propagation (estimated offspring mean) $\hat{\lambda}$ can be compared. More details can be found in [6], [7].
- 3) The interactions between component failures for both the original and simulated cascades are quantified and the probability distribution of the link weights and vertex out-strength and in-strength of the interaction network are compared.
- 4) The spreading capacities of the links quantified from the original and simulated cascades are compared by some defined similarity indices as described below.

For the first and third methods we only need to compare the probability distribution of the total line outages, the link weights, or the vertex out-strength and in-strength for the original and simulated cascades. For the second method the calculation of the average propagation can be found in [6], [7]. Thus in the rest of this section we will only discuss the fourth method in detail.

Let \mathcal{L}_1 , \mathcal{L}_2 , and \mathcal{L}_3 be the set of links shared by the original and simulated cascades and the links only owned by the original and simulated cascades. Denote the index of link l for the original and simulated cascades respectively by I_l^{ori}

and I_l^{sim} . Five similarity indices are defined as follows.

$$S_1 = \frac{\sum_{l \in (\mathcal{L}_1 \cup \mathcal{L}_3)} I_l^{\text{sim}}}{\sum_{l \in (\mathcal{L}_1 \cup \mathcal{L}_2)} I_l^{\text{ori}}} \quad (13)$$

$$S_2 = \frac{\sum_{l \in \mathcal{L}_1} I_l^{\text{ori}}}{\sum_{l \in (\mathcal{L}_1 \cup \mathcal{L}_2)} I_l^{\text{ori}}} \quad (14)$$

$$S_3 = \frac{\sum_{l \in \mathcal{L}_1} I_l^{\text{sim}}}{\sum_{l \in (\mathcal{L}_1 \cup \mathcal{L}_3)} I_l^{\text{sim}}} \quad (15)$$

$$S_4 = \frac{\sum_{l \in \mathcal{L}_1} I_l^{\text{sim}}}{\sum_{l \in \mathcal{L}_1} I_l^{\text{ori}}} \quad (16)$$

$$S_5 = \sum_{l \in \mathcal{L}_1} \left(\frac{I_l^{\text{sim}} + I_l^{\text{ori}}}{\sum_{l \in \mathcal{L}_1} (I_l^{\text{sim}} + I_l^{\text{ori}})} \frac{I_l^{\text{sim}}}{I_l^{\text{ori}}} \right) \quad (17)$$

S_1 is the ratio between the summation of the link weights of the simulated cascades and that of the original cascades. If S_1 is close to 1.0 the links of the original and simulated cascades have almost the same spreading capacity.

S_2 and S_3 are used to indicate if the shared links play the major role among all links for the original and simulated cascades. If they are near 1.0 it means that the shared links dominate and thus the simulated cascades are similar to the original cascades.

S_4 indicate the similarity between the overall spreading capacity of the shared links of the simulated cascades and that of the original cascades. $S_4 \simeq 1$ will suggest that the overall spreading capacity of the shared links for the simulated cascades are close to that of the original cascades.

But even when $S_4 \simeq 1$ it is still possible that the weight of the same link for the original and simulated cascades can be quite different. Thus S_5 is defined to show if the same link is close to each other. When S_5 is near 1.0 it indicates that at least the most important links of the simulated cascades have spreading capacity close to their counterparts for the original cascades.

Note that the similarity indices defined here can not only be used to compare the similarity of the links obtained from original and simulated cascades but can also be used to compare any two sets of links. For example, we can use them to compare the links from different number of original cascades, which can be denoted by $I_l(M_u^1)$ and $I_l(M_u^2)$ respectively for M_u^1 and M_u^2 original cascades. Since I_l depends on the number of cascades that are used to quantify the interactions, $I_l(M_u^1)$ and $I_l(M_u^2)$ should first be normalized before they are compared. One simple way to normalize them is to divide $I_l(M_u^1)$ by M_u^1/M_u^2 or to divide $I_l(M_u^2)$ by M_u^2/M_u^1 .

V. NUMBER OF CASCADES NEEDED

In the above sections we assume there are a total of M cascades and in section IV we use M_u of them to generate

the tripping probability of each component in generation 0 and the interaction matrix. But two questions remain unanswered, which are how many cascades we need to obtain almost all the interactions between cascading outages and how many cascades we need to obtain the dominant interactions that can be used to generate cascades matching well enough with the original cascades. In this section we discuss these two questions and determine the lower bounds M^{min} and M_u^{min} respectively for M and M_u .

A. Determining Lower Bound for M

More original cascades tend to contain more information about the property of cascading failures of a system, or more specifically the interactions between cascading outages of the components. The added information brought from the added cascades will make the number of identified links increase. However, the number of links will not always grow with the increase of the number of cascades but will saturate after the number of cascades is greater than some number M^{min} , which can be determined by gradually increasing the number of cascades, recording the number of identified links, and finding the smallest number of cascades that can lead to the saturated number of links.

Assume there are a total of N_M different M ranging from very small number to very large number, which are denoted by $M_i, i = 1, 2, \dots, N_M$. The number of links for M_i cascades is denoted by $\text{card}(\mathcal{L}(M_i))$ where $\text{card}(\cdot)$ denotes the cardinality of a set, which is a measure of the number of elements of the set.

For $i = 1, 2, \dots, N_M - 2$ we define

$$\sigma_i = \sigma(\text{card}(\mathcal{L})_i) \quad (18)$$

where

$$\text{card}(\mathcal{L})_i = [\text{card}(\mathcal{L}(M_i)) \cdots \text{card}(\mathcal{L}(M_{N_M}))] \quad (19)$$

and $\sigma(\cdot)$ is the standard deviation of a vector.

The σ_i for $i = N_M - 1$ and $i = N_M$ are not calculated since we would like to calculate the standard deviation for at least 3 data points. Very small and slightly fluctuating σ_i can be used to indicate that the number of links begins to saturate after M_i and thus this M_i is determined as M^{min} .

The M^{min} original cascades can guarantee that the accuracy on statistical values of interest is good and thus can provide a reference solution.

B. Determining Lower Bound for M_u

When we only want to obtain the dominant interactions that can be used to generate cascades matching well enough with the original cascades, we do not need as many as M^{min} cascades but only M_u^{min} cascades to make sure that the propagation capacity of the obtained interaction network $\mathcal{G}(\mathcal{C}, \mathcal{L})$ (denoted by $PC^{\mathcal{G}}$) is consistent with that of the original cascades (denoted by PC^{ori}). Here the physical meaning of the propagation capacity is the average value of the number of caused failures in one cascade. In this section we propose a method to determine M_u^{min} .

Since both PC^{ori} and $PC^{\mathcal{G}}$ vary with M_u we denote them by $PC^{\text{ori}}(M_u)$ and $PC^{\mathcal{G}}(M_u)$. $PC^{\text{ori}}(M_u)$ can be directly obtained from the original cascades by calculating the average value of the number of failures in generation one and the following generations as

$$PC^{\text{ori}}(M_u) = \frac{\sum_{m=1}^{M_u} \sum_{g=1}^{\infty} \text{card}(F_g^{(m)})}{M_u}. \quad (20)$$

In section III we define an index I_l for each link $l: i \rightarrow j$ which is the expected value of the number of failures that are caused through link l . In order to get all the failures that are caused through the link $l: i \rightarrow j$ we set the number of failures of its source vertex i as N_i , which is the total number of failures among all the original cascades. Note that N_i contains not only the failures in generation 0, which serve as trigger of cascading failures, but also the failures caused by other component failures. In this section, however, we only would like to calculate the expected value of the number of failures caused through link $l: i \rightarrow j$ by its source vertex i as generation 0 failures. In this case we need to set N_i^s to be $N_{i,0}$, which is the number of failures of component i in generation 0 among all M_u original cascades. Correspondingly the calculated link index is denoted by I_l' to distinguish with I_l in section III for which $N_i^s = N_i$. Then the propagation capacity of the interaction network can be written as

$$PC^{\mathcal{G}}(M_u) = \frac{\sum_{l \in \mathcal{L}} I_l'(M_u)}{M_u}. \quad (21)$$

When M_u is not large enough it is expected that there will be a big mismatch between $PC^{\text{ori}}(M_u)$ and $PC^{\mathcal{G}}(M_u)$, indicating that the quantified interactions between cascading outages cannot well capture the property of the cascading failure propagation. But with the increase of M_u more information will be obtained and thus the mismatch will gradually decrease. Based on this we increase M_u gradually and identify M_u^{min} as the smallest value that satisfies the following condition

$$|\Delta_{PC}(M_u)| \leq \epsilon_{PC} PC^{\text{ori}}(M_u) \quad (22)$$

where $\Delta_{PC}(M_u) = PC^{\mathcal{G}}(M_u) - PC^{\text{ori}}(M_u)$ and ϵ_{PC} is used to determine the acceptable mismatch.

In order to get M_u^{min} we start from very small M_u , such as $M_u^0 = 100$, and calculate the mismatch $\Delta_{PC}(M_u)$. If the condition in (22) is not satisfied M_u is increased by a big step ΔM_1 and recalculate $\Delta_{PC}(M_u)$ with the new M_u ; otherwise M_u is decreased by a small step ΔM_2 until the last M_u for which the condition in (22) is still satisfied.

Note that the number of unnecessary original cascade simulation runs M_{un} is always less than ΔM_1 and actually can be determined by $M_{un} = N_{\Delta M_2} \Delta M_2$ where $N_{\Delta M_2}$ is the number of times ΔM_2 is used for iterations. Also the obtained M_u^{min} is not greater than the smallest possible M_u by ΔM_2 . By decreasing ΔM_1 we can decrease the upper bound of M_{un} but cannot necessarily decrease M_{un} . Differently, by decreasing ΔM_1 we can surely increase the accuracy of the obtained M_u^{min} . But smaller ΔM_1 or ΔM_2 will increase the time for getting the interaction network and quantifying I_l' for

the links. The selection of ΔM_1 and ΔM_2 can be guided by ϵ_{PC} . Smaller mismatch ϵ_{PC} will need greater M_u^{min} and thus larger ΔM_1 and ΔM_2 can be chosen to avoid too many times of calculating the interaction network and quantifying I_l' .

VI. CASCADING FAILURE MITIGATION MEASURES

Since the system-level failures of a complex system are actually caused by the interaction of components, one possible mitigation measure can be preformed by weakening some key links between component failures, which will possibly stop the propagation of cascading failures at an initial stage. Here weakening key links means reducing the corresponding element in the interaction matrix \mathbf{B} .

After validating the proposed interaction model in section IV, this model can be applied to study how the interactions between component failures influence the cascading failure risk and to efficiently validate the effectiveness of the mitigation measures based on the weakening of key links.

In real systems the weakening of key links can be implemented by blocking some specific protective relays. The zone 3 relay blocking method called adaptive distance relay scheme has been discussed in [30]. In this paper relays are blocked under the condition of the tripping of the lines corresponding to the source vertices of the key links. Since the key links can cause tremendous expected number of failures and thus play crucial roles in the propagation of cascading failures it should be beneficial to the overall security of the system to stop the propagation from the source vertices of key links to the destination vertices by blocking the operation of the relay of the destination vertices, thus securing time for the operators to take remedial actions, such as re-dispatching the generation or even shedding some loads, and finally helping mitigate catastrophic failures.

This relay blocking strategy under the condition of some specific line tripping can be considered as a wide area protection scheme, which can be simulated in AC OPA model by adding a relay blocking module. Note that overloaded lines can be tripped not only by zone 3 relay but can also by other causes, such as tree flashover. The tripping of some lines in two significant outages in the western US in 1996 [1], the US-Canada Blackout on August 14, 2003 [2], and the outage in Italy on September 28, 2003 [3] can all be attributed to tree flashover to some extent, which has been discussed in [18] and [22].

In order to simulate the implementation of the mitigation strategy by weakening some links in real systems and also to compare with the results from the interaction model we add a relay blocking module in AC OPA model which decreases the tripping probability of the some overloaded lines and thus simulates the the weakening of key links. When the line corresponding to the source vertex of a key link is tripped and further causes the overloading of the line corresponding to its destination vertex, the destination vertex line will be tripped with a reduced probability to simulate the part of the role played by blocking of its relay in preventing the line tripping. Due to the reduced tripping probability AC OPA will probably go to its next inner iteration without tripping this

destination vertex line and thus AC OPF will be calculated and generation re-dispatching and load shedding will be performed to eliminate the overloading of the destination vertex line.

VII. RESULTS

This section presents results for interaction matrix, interaction network, and interaction model. The cascading outage data is produced by open-loop AC OPA simulation [16], [17] on IEEE 118-bus system, which is standard except that the line flow limits are determined with the same method in [7]. The probability for initial line outage is $p_0 = 0.0001$ and the load variability $\gamma = 1.67$, which are the same as [6], [7].

AC OPA is a variant of the basic OPA [11]–[14]. Basic OPA use DC power flow, for which only active power is considered and the bus voltages are assumed constant. In contrast, AC OPA uses AC power flow and thus can consider reactive power and voltage. The operation mode of the system is first determined by AC OPF and load shedding and will be readjusted by AC OPF until there is no further outage or failure once outages happen. Both reactive power and voltage constraints are taken into account in AC OPF.

For testing the interaction quantifying method and the proposed interaction model, AC OPA simulation at base case load level is run so as to produce 5000 cascading outages with a nonzero number of line outages. In section VII-B–VII-E all 5000 original cascades are used while in section VII-F using smaller number of original cascades is discussed.

Note that we only take generating original cascades by AC OPA on IEEE 118-bus system as an example. AC OPA is only one of the many models that can be used to generate the original cascades needed by the proposed interaction model. We can also generate original cascades with basic OPA, improved OPA [15], or OPA with slow process [18]. The test systems can also be real system models used in previous literature, such as the 568-bus Northeast Power Grid of China [7], [15], [18] or the 1553-bus WECC system [14].

Moreover, the original cascades can also come from utility data, such as the Transmission Availability Data System (TADS), which has 10512 outages recorded by North America utility and can be grouped into 6316 cascades based on outages' timing [8]. Although the TADS data have been successfully used to study the propagation of cascading failures by using the branching process model in [8] it is still an open question whether or not the statistical data from utilities are sufficient for obtaining the interaction matrix and interaction network discussed in this paper. One way to check this is to apply the method proposed in section V-B to determine the minimum number of cascades that are needed to obtain the dominant interactions, which can further be used to generate cascades matching well enough with the original cascades. This definitely deserves more careful discussion in our future work.

Both AC OPA and the interaction model are implemented with Matlab and all tests are carried out on a 3.4 GHz Intel(R) Core(TM) i7-3770 based desktop.

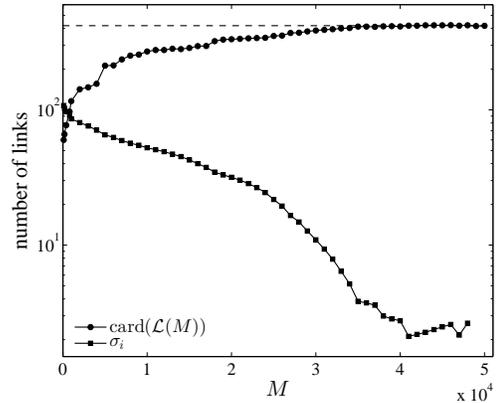


Fig. 4. Number of links for different M . The dash horizontal line indicates $\text{card}(\mathcal{L}(M)) = 419$.

A. Number of Cascades Needed

In this section we determine the lower bounds for M and M_u by using the method in section V. The number of identified links $\text{card}(\mathcal{L}(M))$ for different M ranging from 100 to 50000 is shown in Fig. 4. We can see that the number of links first grows with the increase of M and finally saturate and only fluctuate slightly when M is large enough. There are a total of $N_M = 54$ different M . In Fig. 4 we also show σ_i for $1 \leq i \leq N_M - 2$. It is very clear that σ_i decreases with the increase of M and finally stabilizes at $M = 41000$. Therefore, we choose $M^{\min} = 41000$ for which the number of identified links is 419. The largest number of links for $\text{card}(\mathcal{L}(M_i)), i = 1, 2, \dots, N_M$ is 423. The number of identified links for $M^{\min} = 41000$ is greater than 99% of the largest number of links. In the rest of this paper we will simulate 41000 cascades for both the AC OPA and the interaction simulation.

In order to determine M_u^{\min} we choose $M_u^0, \epsilon_{PC}, \Delta M_1$, and ΔM_2 in section V-B as 100, 0.01, 1000, and 100. By setting ΔM_2 to be 100 we can guarantee that the obtained M_u^{\min} is not greater than the smallest possible M_u by 100. After 9 ΔM_1 iterations and 1 ΔM_2 iteration M_u^{\min} is determined as 8000, which accounts for 19.51% of M^{\min} . The number of unnecessary original cascade simulation runs M_{un} in this case is only 100. If we set ϵ_{PC} as 0.05 and correspondingly decrease ΔM_1 and ΔM_2 to 200 and 10, M_u^{\min} will be determined as 3680, which only accounts for 8.98% of M^{\min} , after 19 ΔM_1 iterations and 2 ΔM_2 iterations. The number of unnecessary original cascade simulation runs M_{un} in this case is only 20.

When we have already generated a large number of cascades we can show how the propagation capacity obtained from original cascades and the interaction network changes with the increase of M_u . This is shown in Fig. 5, in which we can clearly see the trend of the decreasing mismatch between $PC^{\text{ori}}(M_u)$ and $PC^{\mathcal{G}}(M_u)$.

In Fig. 5 we can also see that the quantified interaction network tends to obtain an overestimated propagation capacity for a small number of M_u . As has been tentatively discussed in section II, this is because for any component j no component

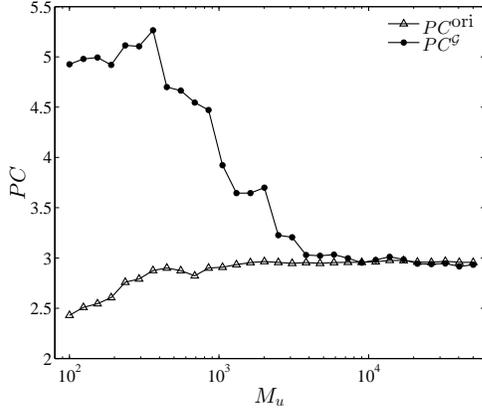
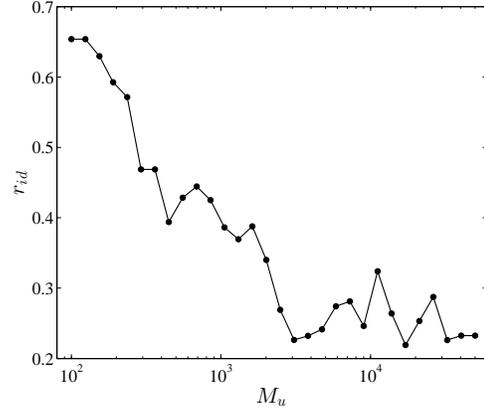
Fig. 5. Propagation capacity for different M_u .

Fig. 6. Ratio of completely cause-indistinguishable components.

has much greater a_{ij} than the others when only using a small number of M_u and thus several components have to be determined as the cause. This further leads to the overestimation of the propagation of cascading failures.

In order to show this we analyze how \mathbf{A} changes with the increase of M_u in more detail. Let $\mathcal{C}^{\text{caused}}$ denote the set of components that are caused by other components in \mathbf{A} , which actually corresponds to the columns of \mathbf{A} with nonzero elements. For any $j \in \mathcal{C}^{\text{caused}}$, if $\text{card}(\{i|a_{ij} > 0\}) > 1$ and $\max_{i \in \mathcal{C}} a_{ij} = \bar{a}_j$ where \mathcal{C} is all the components and \bar{a}_j is the average value for all $a_{ij} > 0$ then component j is called completely cause-indistinguishable. The set of completely cause-indistinguishable components is denoted by \mathcal{C}^{id} . Actually when the given condition holds it is easy to prove by contradiction that all nonzero a_{ij} will be equal. Thus in this case it is completely impossible to distinguish which cause of component j is more possible than the others. The ratio of completely cause-indistinguishable components among all caused components can be calculated as

$$r_{id} = \frac{\text{card}(\mathcal{C}^{\text{id}})}{\text{card}(\mathcal{C}^{\text{caused}})}. \quad (23)$$

We show r_{id} in Fig. 6 and it is clearly seen that the completely cause-indistinguishable components account for a large proportion when M_u is small and will gradually decrease to a relatively low level with the increase of M_u . The high ratio of the completely cause-indistinguishable components for a small number of M_u leads to the overestimated interactions between component failures and further explains the too great propagation capacity of the interaction network.

B. Interaction Matrix and Interaction Network

There are 186 lines in IEEE 118-bus system and thus \mathbf{B} is a 186×186 square matrix. Table I shows the number of components n , the number of cascades used to quantify the interactions M_u , the number of links $\text{card}(\mathcal{L})$, which is also the number of \mathbf{B} 's nonzero elements, and the ratio of nonzero elements $r = \text{card}(\mathcal{L})/n^2$. It is seen that r is very small, indicating that the interaction matrix is very sparse and that only a small fraction of lines interact with each other.

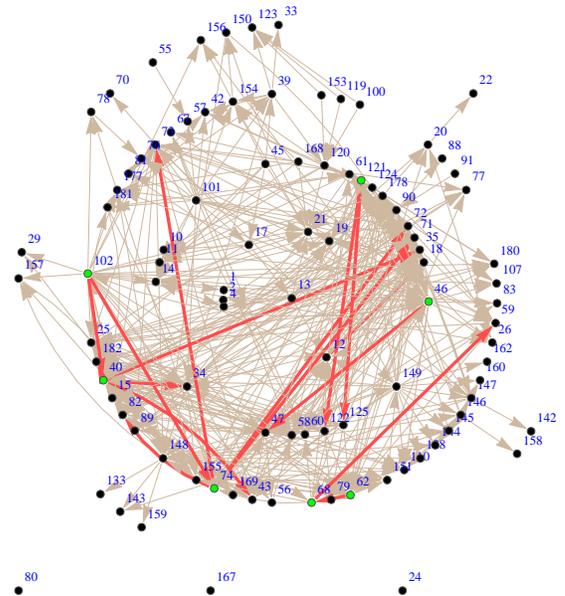


Fig. 7. Interaction network for using 41000 original cascades.

TABLE I
NONZERO ELEMENTS IN \mathbf{B} FOR IEEE 118-BUS SYSTEM

model	n	M_u	$\text{card}(\mathcal{L})$	r
AC OPA	186	41000	419	0.0121
AC OPA	186	8000	252	0.00730
AC OPA	186	3680	156	0.00450

The corresponding directed weighted interaction network is shown in Figs. 7–9, in which the dots denote lines in IEEE 118-bus system and the arrows denote the links between lines. Here we do not show the weights of the links but only the topology of the interaction network. This network is different from the one-line diagram of IEEE 118-bus system, for which the vertices are buses and the undirected links between vertices are lines.

We compare the links obtained from 8000 and 3680 original cascades with the reference 41000 original cascades by calculating the similarity indices defined in section IV-B, which are shown in Table II. As is discussed in section IV-B, since I_l depends on the number of cascades that are used the index

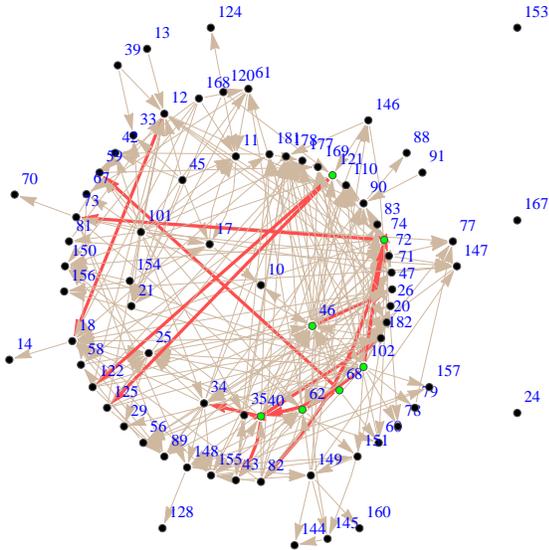


Fig. 8. Interaction network for using 8000 original cascades.

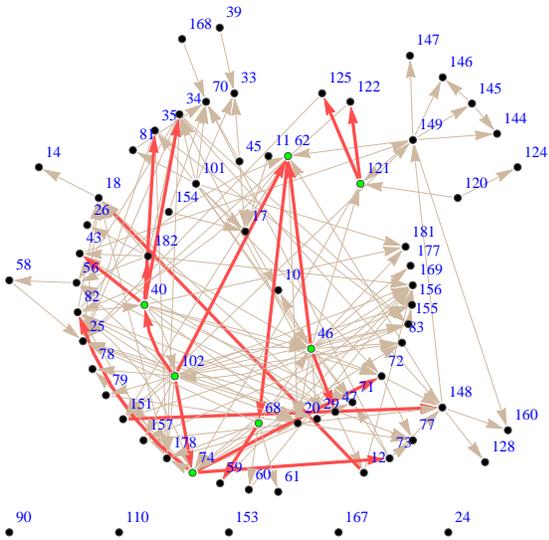


Fig. 9. Interaction network for using 3680 original cascades.

I_l of the links should first be normalized before calculating the similarity indices. Specifically, we divide the I_l for 41000 cascades by $41000/8000$ and $41000/3680$. Results in Table II shows that the first four indices are all close to 1 and thus indicate the links from smaller number of cascades are similar to those for more cascades. But the fifth index S_5 , which requires more strict similarity, is not so close to 1, indicating that some information is missing for fewer original cascades. However, we will show in section VII-D that the links obtained from simulated more cascades with the proposed interaction model can be very similar to those from 41000 original cascades.

TABLE II
SIMILARITY INDICES FOR 8000 AND 3680 ORIGINAL CASCADES
COMPARED WITH 41000 ORIGINAL CASCADES

model	M_u	S_1	S_2	S_3	S_4	S_5
AC OPA	8000	1.02	0.996	0.991	1.02	2.79
AC OPA	3680	1.04	0.987	0.985	1.04	6.41

C. Key Link and Key Component Identification

In this section key links and key components that play important roles in the propagation of cascading failures are identified by using the method in section III. Both ϵ_l and ϵ_s are chosen as 0.15 (slightly greater than $1/10$) to make the weights of all key links and all key components are in the same order.

The identified key links, which are actually line pairs in IEEE 118-bus system, and their weights I_l for the three cases separately using 41000, 8000, and 3680 original cascades are listed in Table III. They are also shown in Figs. 7–9 by red arrows. The numbers in the parentheses for $I_l(8000)$ and $I_l(3680)$ are the ranking of the key links. We can see that the identified key links for the three cases are almost the same and the ranking of the links is also quite similar. The link $(53, 54) \rightarrow (49, 51)$ is only identified for using 41000 cascades and the links $(65, 66) \rightarrow (45, 46)$, $(35, 36) \rightarrow (45, 46)$, and $(80, 97) \rightarrow (80, 96)$ are only identified for using 3680 cascades. Although the link $(53, 54) \rightarrow (49, 51)$ is not identified for using 8000 and 3680 cascades, the corresponding I_l for the two cases separately rank the 16th and 19th and are equal to 378 and 157, which are much greater than the average values of all link indices 109 and 82. For the three cases the number of key links are only 3.82%, 5.95%, and 11.54% of all the links but the summation of their weights accounts for 88.08%, 85.12%, and 87.65% of the total weights of all links.

The identified key components, the corresponding lines, and their out-strengths for using using 41000, 8000, and 3680 original cascades are listed in Table IV, in which the numbers in the parentheses are the ranking of the key components. It is seen that the identified key components and their ranking are exactly the same for the three cases. They are also highlighted in Figs. 7–9 by green vertices. The tripping of these lines will cause severe consequences and thus should be prevented to the greatest extent. For IEEE 118-bus system there are a total of 186 components and among them 97, 76, 68 components are involved in the 41000, 8000, and 3680 original cascades. For the three number of original cascades, the number of key components are 3.76% of all components and separately 7.22%, 9.21%, and 10.29% of the involved components and the summation of the out-strengths of the key components accounts for 89.12%, 88.83%, and 88.65% of the total out-strengths of the involved components.

The identified key links and key components are denoted on the one-line diagram of IEEE 118-bus system, which is shown in Fig. 10. It is seen that the lines corresponding to the source and destination vertices of some key links can be topologically far away from each other, such as link $(65, 66) \rightarrow (53, 54)$ and $(65, 66) \rightarrow (29, 31)$, although for most key links the source and destination vertices are lines that are topologically close to each other, such as link $(53, 54) \rightarrow (51, 52)$ and

TABLE III
KEY LINKS FOR IEEE 118-BUS SYSTEM

$i \rightarrow j$	line pairs	$I_l(41000)$	$I_l(8000)$	$I_l(3680)$
74 → 72	(53, 54) → (51, 52)	12582	2486 (2)	1150 (1)
74 → 73	(53, 54) → (52, 53)	12469	2535 (1)	1150 (2)
40 → 34	(29, 31) → (27, 28)	11920	2305 (3)	1031 (3)
40 → 35	(29, 31) → (28, 29)	11421	2167 (4)	974 (5)
74 → 82	(53, 54) → (56, 58)	10802	2153 (5)	1000 (4)
62 → 68	(45, 46) → (45, 49)	9865	1915 (6)	940 (6)
121 → 122	(77, 78) → (78, 79)	9601	1912 (7)	862 (7)
121 → 125	(77, 78) → (79, 80)	9599	1912 (8)	862 (8)
40 → 182	(29, 31) → (114, 115)	6687	1261 (9)	592 (9)
46 → 47	(35, 36) → (35, 37)	5536	1092 (10)	502 (10)
12 → 18	(11, 12) → (13, 15)	5388	1001 (11)	450 (11)
40 → 43	(29, 31) → (27, 32)	4475	886 (12)	404 (12)
68 → 59	(45, 49) → (43, 44)	3690	706 (13)	347 (13)
102 → 74	(65, 66) → (53, 54)	3135	635 (14)	273 (14)
74 → 71	(53, 54) → (49, 51)	1977	-	-
102 → 40	(65, 66) → (29, 31)	1968	415 (15)	178 (15)
102 → 62	(65, 66) → (45, 46)	-	-	178 (16)
46 → 62	(35, 36) → (45, 46)	-	-	178 (17)
151 → 148	(80, 97) → (80, 96)	-	-	175 (18)

TABLE IV
KEY COMPONENTS FOR IEEE 118-BUS SYSTEM

key component	line	$s_i^{\text{out}}(41000)$	$s_i^{\text{out}}(8000)$	$s_i^{\text{out}}(3680)$
74	(53, 54)	37925	7577 (1)	3467 (1)
40	(29, 31)	34613	6651 (2)	3015 (2)
121	(77, 78)	19219	3833 (3)	1728 (3)
62	(45, 46)	9892	1926 (4)	944 (4)
102	(65, 66)	8039	1826 (5)	892 (5)
46	(35, 36)	6645	1377 (6)	734 (6)
68	(45, 49)	6210	1211 (7)	594 (7)

(53, 54) → (52, 53). This is because the interactions and links are obtained from simulated cascades generated by AC OPA model which not only considers the topology of the power network but also other physics of the system, such as power flow and the operator response. These factors can also make

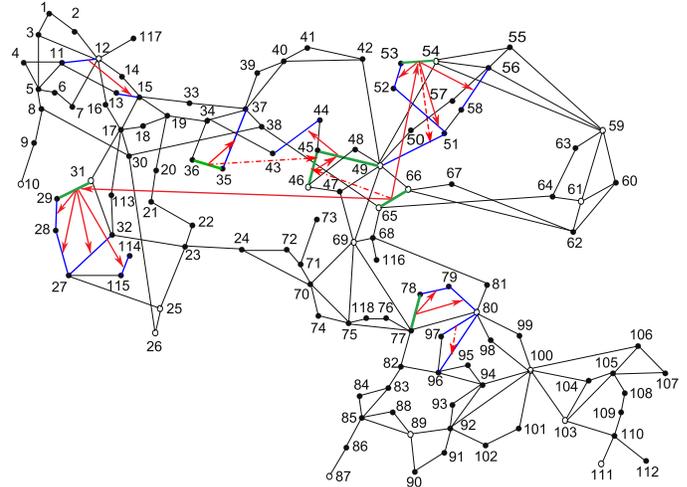


Fig. 10. Key links and key components for IEEE 118-bus system. Key links shared by the three cases for using 41000, 8000, and 3680 cascades are denoted by red solid arrow; key links only for using 41000 cascades are denoted by red dash arrow; key links only for using 3680 cascades are denoted by red dash-dotted arrow; key components are denoted by green lines and the other lines involved in the key links are denoted by blue lines.

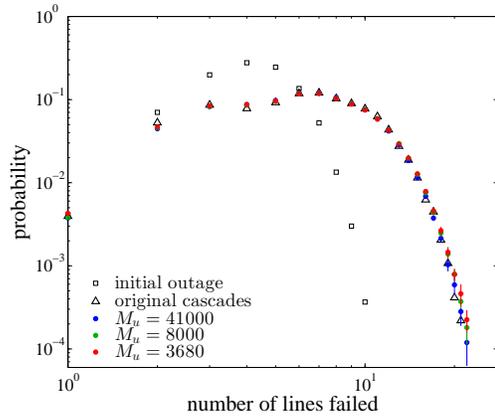


Fig. 11. Probability distributions of the total number of line outages for original and simulated cascades. Triangles indicate total numbers of line outages of the original cascades; blue, green, and red dots separately indicate total numbers of line outages of the simulated cascades for using 41000, 8000, and 3680 original cascades to quantify interactions; vertical lines indicate the standard deviations; squares indicate initial line outages.

some components tightly coupled.

D. Model Validation

In this section the proposed interaction model is validated with the four methods discussed in section IV-B. The probability distributions of the total number of line outages for original cascades (triangles) and simulated cascades (dots) by using 41000, 8000, and 3680 original cascades to quantify interactions are shown in Fig. 11. For the simulated cascades we simulate 41000 cascades for 20 times for each case and show their average probability distribution and the standard deviations (vertical lines).

It is seen that the distributions of total line outages of the original and simulated cascades match well and the standard

deviations of the probability distributions of the simulated cascades are small. This suggests that the proposed interaction model can generate cascades with similar statistical properties to the original cascades. The dramatic difference between the distributions of the initial and total outages also suggest that the cascading failure is able to propagate a lot. This is because of the interaction between component failures denoted by the sparse interaction matrix. If all elements of \mathbf{B} are zero and the components do not interact at all, all cascades will stop immediately after initial line outages and the distribution of the total line outages will be the same as initial line outages. Thus, although being sparse, the interaction matrix does take effect.

To quantitatively compare the original cascades and the simulated cascades, the branching process is applied to estimate their average propagation. The average value $m_{\hat{\lambda}}$ and the standard deviation $\sigma_{\hat{\lambda}}$ of the average propagation for 20 times of simulation are listed in Table V, in which M_u is the number of original cascades used to quantify the interactions between component failures and M is the number of cascades simulated by the interaction model. It is seen that the estimated average propagation of the simulated cascades are very close to that of the original cascades and the standard deviations are very small, indicating that the simulated cascades have similar propagation property to the original cascades.

TABLE V
AVERAGE PROPAGATION FOR ORIGINAL AND SIMULATED
CASCADES

model	M_u	M	$m_{\hat{\lambda}}$	$\sigma_{\hat{\lambda}}$
AC OPA	–	41000	0.402	–
interaction	41000	41000	0.402	0.000969
interaction	8000	41000	0.410	0.000800
interaction	3680	41000	0.412	0.000899

The complementary cumulative distributions (CCD) of the link weights for the original and simulated cascades for $M_u = 41000$, $M_u = 8000$, and $M_u = 3680$ are shown in Fig. 12. The CCD for simulated cascades are the distribution for all the links obtained from 20 times of simulation. Note that we group the links with very big weights together and calculate the CCD for their average value to avoid the possible unreliable estimation for rarer events since the number for each of them can be very small. Also when we calculate the CCD the zero elements are considered as links with zero weights. Fig. 12 shows that the two distributions match very well. Both of them follow obvious power law and can range from 1 to more than 10000, suggesting that a small number of links can cause much greater consequences than most of the others.

The CCD of the vertex out-strength and in-strength for original and simulated cascades are shown in Figs. 13–14. Similar to link weights, the CCD for simulated cascades are the distribution for all the vertex out-strength and in-strength obtained from 20 times of simulation. We also group the vertices with very big out-strength or in-strength together and the components that do not appear in the interaction network are considered as vertices with zero out-strength and in-strength. The strength distributions of the original and sim-

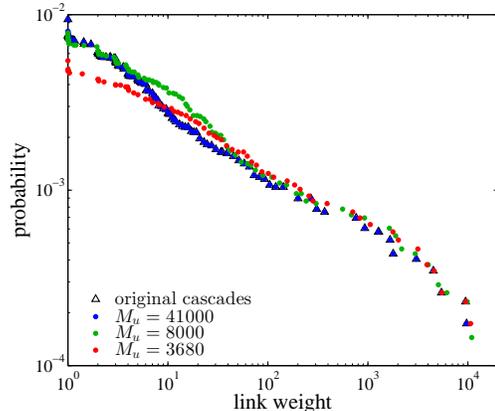


Fig. 12. CCD of the link weight for original and simulated cascades. Triangles indicate CCD of the link weight of original cascades; blue, green, and red dots separately indicate CCD of the link weight of the simulated cascades for using 41000, 8000, and 3680 original cascades to quantify interactions.

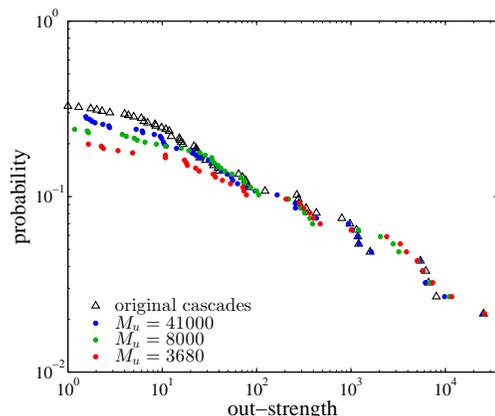


Fig. 13. CCD of the out-strength for original and simulated cascades.

ulated cascades match very well, indicating that the simulated cascades share similar features to the original cascades from an overall point of view. An obvious power law behavior can also be seen, which means that the failure of most vertices (components) have small consequences while a small number of them have much greater impact.

The five similarity indices defined in section IV-B for the quantified links from the simulated 41000 cascades and the 41000 original cascades are listed in Table VI. It is seen that all five indices are close to 1.0 and thus the links obtained from the original and simulated cascades are actually quite similar. The standard deviations of the similarity indices for 20 times of simulation are listed in the parentheses and are all very small. The similarity indices between the quantified links from 8000 and 3680 original cascades and those from 41000 cascades are listed in Table II. By comparing the results in Table II and Table VI we can see that the first four indices are only slightly different but the fifth index significantly improves for simulating 41000 cascades with the interaction model by using fewer original cascades to quantify interactions, indicating that the missing information for fewer original cascades can be recovered to a great extent

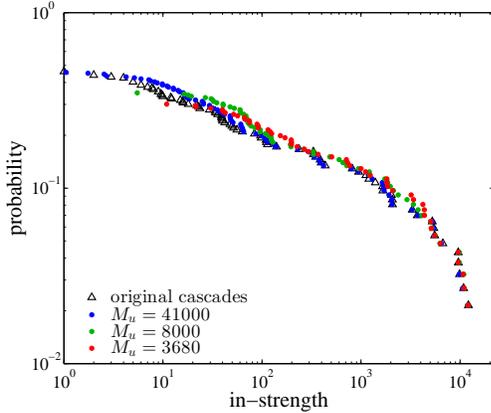


Fig. 14. CCD of the in-strength for original and simulated cascades.

by simulating more cascades with the interaction model, which has high-level probabilistic property and can reveal more rare events by doing a large number of simulations. It is also not surprising that using fewer original cascades cannot get as good results as using the whole 41000 original cascades since fewer original cascades will inevitably miss some information to some extent.

TABLE VI
SIMILARITY INDICES FOR ORIGINAL AND SIMULATED CASCADES

M_u	S_1	S_2	S_3	S_4	S_5
41000	0.994 (0.00900)	0.994 (0.000619)	0.973 (0.0168)	0.974 (0.0194)	1.01 (0.0122)
8000	1.05 (0.00570)	0.989 (0.00280)	0.973 (0.00930)	1.04 (0.0146)	1.14 (0.0379)
3680	1.07 (0.00540)	0.987 (0.00240)	0.944 (0.00180)	1.03 (0.00690)	1.17 (0.0207)

E. Cascading Failure Mitigation

We assume that 90% of the tripping of overloaded lines are due to the operation of zone 3 relays. For the interaction model the weakening of the key links is simulated by reducing the corresponding elements in the interaction matrix by 90%. For each of the three cases in which $M_u = 41000$, $M_u = 8000$, and $M_u = 3680$, the key links identified in Table III of section VII-C are weakened by reducing the corresponding elements in $\mathbf{B}(M_u)$ by 90%. By doing this we get $\mathbf{B}_{\text{int}}(M_u)$. For comparison the same number of randomly chosen links are also weakened in the same way, for which $\mathbf{B}_{\text{rand}}(M_u)$ is obtained. Cascading failures are separately simulated with the proposed model by using $\mathbf{B}_{\text{int}}(M_u)$ and $\mathbf{B}_{\text{rand}}(M_u)$. The two mitigation strategies are respectively called intentional mitigation and random mitigation. Each case for each mitigation strategy is simulated for 20 times.

Fig. 15 shows the probability distributions of total line outages under the two mitigation strategies for $M_u = 41000$, $M_u = 8000$, and $M_u = 3680$. It is seen that the risk of large-scale cascading failures can be significantly mitigated by weakening only a small number of key links. By contrast, the mitigation effect is minor if the same number of randomly chosen links are weakened.

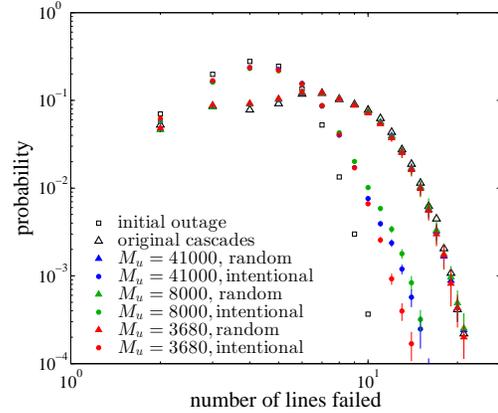


Fig. 15. Probability distributions of the total number of line outages under two mitigation strategies. Triangles indicate total numbers of line outages of the original cascades under no mitigation; blue, green, and red triangles and dots separately indicate total numbers of line outages of the simulated cascades under random and intentional mitigation for using 41000, 8000, and 3680 original cascades to quantify interactions; vertical lines indicate the standard deviations; squares indicate initial line outages.

This can be explained by the power law distribution of the link weights. Most links have small weights and only a small number of links have much greater weights. When randomly weakening links it is more possible to choose small-weight links and thus the total weights of the weakened links for random mitigation can be significantly weaker than that for the intentional mitigation. The probability for exactly choosing the key links and thus getting the largest possible total link weights for $M_u = 41000$, $M_u = 800$, and $M_u = 3680$ are respectively as low as 3.10×10^{-29} , 1.91×10^{-24} , and 5.93×10^{-24} . The summation of the weights of weakened links for intentional and random mitigation are listed in Table VII, in which $m_{I_{\mathcal{L}^{\text{key}}}}$ and $\sigma_{I_{\mathcal{L}^{\text{key}}}}$ separately denote the average value and standard deviation of the total weights of the weakened links. We can see that the total weights of the weakened links for intentional mitigation is more than a order greater than those for random mitigation. The standard deviations of the total weights of the weakened links for random mitigation are big because the distribution of the link weights follows power law and the link weights can vary significantly. The big standard deviations cannot be decreased by doing a large number times of simulation, which indicates that the random mitigation strategy is not stable and the mitigation effects between different random mitigation can be quite different.

TABLE VII
LINK WEIGHTS FOR DIFFERENT MITIGATION STRATEGIES

model	mitigation strategy	M_u	$m_{I_{\mathcal{L}^{\text{key}}}}$	$\sigma_{I_{\mathcal{L}^{\text{key}}}}$
interaction	intentional	41000	121114	0
interaction	random	41000	6126	6624
interaction	intentional	8000	23380	0
interaction	random	8000	1737	1570
interaction	intentional	3680	11247	0
interaction	random	3680	1043	696

To quantitatively compare the effects of different mitigation

measures, the branching process is applied to estimate the average propagation of the original and simulated cascades under two mitigation strategies. The average value and standard deviation of the estimated average propagation are listed in Table VIII. It is seen that the average propagation decreases dramatically under intentional mitigation while decreases only a little under random mitigation. Also the relative standard deviations of the average propagation for random mitigation are much higher than those for intentional mitigation, which can be explained by the big standard deviation of the total weights of the weakened links for random mitigation.

TABLE VIII
AVERAGE PROPAGATION FOR DIFFERENT MITIGATION STRATEGIES

model	mitigation strategy	M_u	M	$m_{\hat{\lambda}}$	$\sigma_{\hat{\lambda}}$
interaction	intentional	41000	41000	0.0966	0.000657
interaction	random	41000	41000	0.391	0.0128
interaction	intentional	8000	41000	0.113	0.00110
interaction	random	8000	41000	0.394	0.0154
interaction	intentional	3680	41000	0.0965	0.000711
interaction	random	3680	41000	0.391	0.0143

In order to simulate the implementation of the mitigation strategy by weakening some links in real systems we add a relay blocking module in AC OPA model. For intentional or random mitigation, when the source vertices of the predetermined links fail and the destination vertices of corresponding links become overloaded and will be tripped by protective relays, the probability of the operation of the relays will be reduced by 90% and AC OPA will have a much greater chance to go to the next inner iteration without tripping this overloaded destination vertex line, in which AC OPF is performed to simulate the re-dispatching of generation and some loads are shed if necessary in order to eliminate the violation of the line limits. In this way the AC OPA simulations can get cascades under mitigation strategies.

In Fig. 16 we compare the average probability distributions of the total numbers of the line outages for AC OPA model and interaction model under intentional mitigation for $M_u = 41000$, $M_u = 8000$, and $M_u = 3680$. Note that the identified key links for $M_u = 41000$, $M_u = 8000$, and $M_u = 3680$ are separately denoted by $\mathcal{L}^{\text{key}}(41000)$, $\mathcal{L}^{\text{key}}(8000)$, and $\mathcal{L}^{\text{key}}(3680)$ and the AC OPA model simulation is performed for three times, each of which respectively weakens $\mathcal{L}^{\text{key}}(41000)$, $\mathcal{L}^{\text{key}}(8000)$, and $\mathcal{L}^{\text{key}}(3680)$. It is seen that for all three cases using different M_u the distributions for both models match very well under the intentional strategies. Results for random mitigation are similar and thus are not given.

The branching process is also applied to estimate the average propagation of the AC OPA and interaction model under intentional mitigation strategy, which are listed in Table IX. The average propagations for two models match each other very well and the standard deviations $\sigma_{\hat{\lambda}}$ of the average propagation are very small, indicating that first quantify interaction by using fewer original cascades (such as 8000 or 3680 cascades) and then performing the interaction model simulation to get 41000 cascades can get consistent and almost

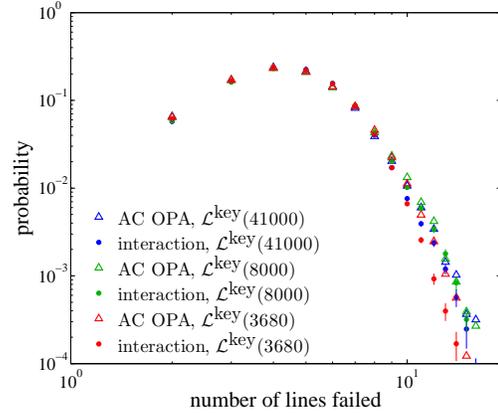


Fig. 16. Probability distributions of the total number of line outages under intentional mitigation for AC OPA and interaction model simulations. Blue, green, and red triangles indicate total numbers of line outages for AC OPA model simulation; blue, green, and red dots indicate total number of line outages for simulated cascades using 41000, 8000, and 3680 original cascades to quantify interactions; vertical lines indicate the standard deviations.

TABLE IX
AVERAGE PROPAGATION FOR AC OPA AND INTERACTION MODEL

model	mitigation strategy	M_u	M	$m_{\hat{\lambda}}$	$\sigma_{\hat{\lambda}}$
AC OPA	intentional ($\mathcal{L}^{\text{key}}(41000)$)	–	41000	0.0944	–
interaction	intentional ($\mathcal{L}^{\text{key}}(41000)$)	41000	41000	0.0966	0.000657
AC OPA	intentional ($\mathcal{L}^{\text{key}}(8000)$)	–	41000	0.109	–
interaction	intentional ($\mathcal{L}^{\text{key}}(8000)$)	8000	41000	0.113	0.00110
AC OPA	intentional ($\mathcal{L}^{\text{key}}(3680)$)	–	41000	0.0998	–
interaction	intentional ($\mathcal{L}^{\text{key}}(3680)$)	3680	41000	0.0965	0.000711

the same results as directly simulating 41000 cascades with the AC OPA model.

F. Efficiency

In this section the improvement of efficiency brought about by the proposed interaction model is discussed. The timing for simulating 41000 cascades for AC OPA simulation and the interaction model simulation based on different number of original cascades M_u is shown in Table X, in which T_1 , T , and T_2 are respectively the time for AC OPA simulation, for calculating the probability that each component fails in generation 0 and the interaction matrix, and interaction model simulation. We can see that it is more time efficient to first quantify the interactions between the component failures with $M_u \ll M$ original cascades and then perform the interaction model simulation than it is to directly simulate M cascades with AC OPA model.

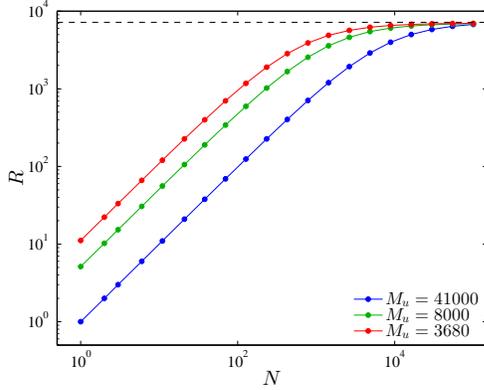


Fig. 17. R for different N . Blue, green, and red dots indicate the efficiency improvement R for using 41000, 8000, and 3680 original cascades. The dash horizontal line indicates $R = 7200$.

TABLE X
EFFICIENCY IMPROVEMENT OF INTERACTION MODEL

model	M_u	M	T_1 (hour)	T (second)	T_2 (second)
AC OPA	–	41000	57	0	0
interaction	41000	41000	57	95	29
interaction	8000	41000	11	15	29
interaction	3680	41000	5	7	29

When many times of simulation need to be performed the advantage of first quantifying the interactions and then doing the interaction model simulation will become more obvious. Assume it takes t_1 and t_2 to generate one cascade from AC OPA model and the interaction model and we have $t_1 \gg t_2$. As in Table X, the time for calculating the probability that each component fails in generation 0 and the interaction matrix by using M_u cascades is denoted by $T(M_u)$. To get N sets of M cascades, the ratio between the simulation time for AC OPA model and that for first quantifying the interactions and then performing interaction model simulation is

$$R = \frac{NMt_1}{M_u t_1 + T(M_u) + NMt_2} \quad (24)$$

and by letting $NM \rightarrow \infty$ we can get

$$\lim_{NM \rightarrow \infty} R = \lim_{NM \rightarrow \infty} \frac{t_1}{\frac{M_u t_1}{NM} + \frac{T(M_u)}{NM} + t_2} = \frac{t_1}{t_2} \quad (25)$$

which indicate a significant efficiency improvement for the interaction model simulations compared with the AC OPA simulations.

In our case $t_1 \simeq 5.04s$ and $t_2 \simeq 0.0007s$. Thus $\lim_{NM \rightarrow \infty} R \simeq 7200$. By letting $M = M^{\min} = 41000$, we show how R changes with N for $M_u = 41000$, $M_u = 8000$, and $M_u = 3680$ in Fig. 17, in which we can see that R first quickly increases and then finally saturates at about 7200. Also as expected, when N is not large enough the efficiency improvement for $M_u = 3680$ is better than that for $M_u = 8000$ because smaller M_u will lead to shorter time for obtaining original cascades from AC OPA simulation and also shorter time for quantifying the interactions.

The efficiency improvement can be reflected in studying the effects of mitigation measures. The interaction model

can generate cascades and study the influence of component interactions on cascading failure risk much more time efficiently while reserving most of the general properties of the cascades. In section VII-E we have shown that the simulated cascades from interaction model are consistent with the AC OPA simulation under the same mitigation strategy. However, in order to obtain the cascades under a specific mitigation strategy AC OPA simulation will need about 57 hours to obtain 41000 cascades while the interaction model simulation only requires about 29 seconds to obtain the same number of cascades by simply changing some elements in the already obtained interaction matrix. This efficiency improvement is important if we would like to quickly find out the impact of a mitigation measure.

Further, as mentioned in section IV, the proposed interaction model can be used for online decision-making support by fast predicting the consequences of the events happening in the system. For online operation there is no enough time for performing detailed cascading failure simulations. But we can obtain the interaction matrix offline from simulations of detailed cascading failure models or from statistical utility data and then apply the interaction model to quickly find out the components or areas of the system will most probably be affected so that a fast response can be performed to pull the system back to normal conditions and to avoid or at least reduce the economic and social losses. The efficiency improvement of the interaction model reflected in this aspect is a great advantage compared with many other cascading failure models.

VIII. CONCLUSION

In this paper we quantify the interaction between component failures by calculating the probability that one component failure causes another and obtain the interaction matrix and interaction network. Key links and key components are identified and an interaction model is proposed to simulate cascading failures and study how interactions between component failures influence cascading failure risk.

The interaction quantifying method and interaction model are validated to be able to capture general properties of the original cascades. It is much more time efficient to first quantify the interactions between the component failures with fewer original cascades from more detailed cascading failure model, such as AC OPA, and then perform the interaction model simulation than it is to directly simulate a large number of cascades with a more detailed model.

An obvious power law is found in distributions of the link weights and the vertex out-strength and in-strength, suggesting that a small number of links and components are much more crucial than the others. Cascading failure risks can be greatly mitigated by weakening a few key links, which can be implemented in real systems by wide area protection that blocks the operation of relays of the lines corresponding to the destination vertices of key links when the lines corresponding to the source vertices are tripped.

Further, the proposed interaction model can also be used for online decision-making support by predicting the consequences of the events based on the interaction matrix obtained

offline from simulations of more detailed cascading failure model or from statistical utility data.

ACKNOWLEDGMENT

We gratefully thank anonymous reviewers for their insightful advice that helped greatly improve our paper.

REFERENCES

- [1] NERC (North America Electric Reliability Council), "1996 System Disturbances," (Available from NERC, Princeton Forrestal Village, 116–390 Village Boulevard, Princeton, New Jersey), 2002.
- [2] U.S.-Canada Power System Outage Task Force, "Final report on the August 14th blackout in the United States and Canada," Apr. 2004.
- [3] Union for the Co-ordination of Electricity Transmission (UCTE), "Interim Report of the Investigation Committee on the 28 September 2003 Blackout in Italy," 2003.
- [4] IEEE PES CAMS Task Force on Cascading Failure, "Initial review of methods for cascading failure analysis in electric power transmission systems," *IEEE PES General Meeting*, Pittsburgh PA USA, Jul. 2008.
- [5] I. Dobson, B. A. Carreras, and D. E. Newman, "A loading dependent model of probabilistic cascading failure," *Probability in the Engineering and Information Sciences*, vol. 19, pp. 15–32, Jan. 2005.
- [6] 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 Analysis*, vol. 30, pp. 650–662, Apr. 2010.
- [7] J. Qi, I. Dobson, and S. Mei, "Towards estimating the statistics of simulated cascades of outages with branching processes," *IEEE Trans. Power Systems*, vol. 28, pp. 3410–3419, Aug. 2013.
- [8] I. Dobson, "Estimating the propagation and extent of cascading line outages from utility data with a branching process," *IEEE Trans. Power Systems*, Vol. 27, pp. 2146–2155, Nov. 2012.
- [9] A. G. Phadke and J. S. Thorp, "Expose hidden failures to prevent cascading outages," *IEEE Comput. Appl. Power*, vol. 9, pp. 20–23, 1996.
- [10] 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, pp. 318–326, May 2005.
- [11] I. Dobson, B. A. Carreras, V. E. Lynch, and D. E. Newman, "An initial model for complex dynamics in electric power system blackouts," *34th Hawaii Intl. Conference on System Sciences*, HI, pp. 710–718, Jan. 2001.
- [12] B. A. Carreras, V. E. Lynch, I. Dobson, D.E. Newman, "Critical points and transitions in an electric power transmission model for cascading failure blackouts," *Chaos*, vol. 12, pp. 985–994, Dec. 2002.
- [13] H. Ren, I. Dobson, and B. A. Carreras, "Long-term effect of the n-1 criterion on cascading line outages in an evolving power transmission grid," *IEEE Trans. Power Systems*, vol. 23, pp. 1217–1225, Aug. 2008.
- [14] B. A. Carreras, D. E. Newman, I. Dobson, and N. S. Degala, "Validating OPA with WECC data," *46th Hawaii Intl. Conference on System Sciences*, HI, Jan. 2013.
- [15] S. Mei, F. He, X. Zhang, S. Wu, and G. Wang, "An improved OPA model and blackout risk assessment," *IEEE Trans. on Power Systems*, vol. 24, pp. 814–823, May. 2009.
- [16] S. Mei, Yadana, X. Weng, and A. Xue, "Blackout model based on OPF and its self-organized criticality," *Proceedings of the 25th Chinese Control Conference*, pp. 7–11, 2006.
- [17] S. Mei, Y. Ni, Weng, G. Wang, and S. Wu, "A study of self-organized criticality of power system under cascading failures based on AC-OPA with voltage stability margin," *IEEE Trans. Power Systems*, vol. 23, pp. 1719–1726, Nov. 2008.
- [18] J. Qi, S. Mei, and F. Liu, "Blackout model considering slow process," *IEEE Trans. Power Systems*, vol. 28, pp. 3274–3282, Aug. 2013.
- [19] M. A. Rios, D. S. Kirschen, D. Jawayeera, D. P. Nedic, and R. N. Allan, "Value of security: modeling time-dependent phenomena and weather conditions," *IEEE Trans. Power Systems*, vol. 17, pp. 543–548, Aug. 2002.
- [20] D. S. Kirschen, D. Jawayeera, D. P. Nedic, and R. N. Allan, "A probabilistic indicator of system stress," *IEEE Trans. Power Systems*, vol. 19, pp. 1650–1657, Aug. 2004.
- [21] M. Anghel, K. A. Werley, and A. E. Mottor, "Stochastic model for power grid dynamics," *40th Hawaii Intl. Conference on System Sciences*, Hawaii, HI, Jan. 2007.
- [22] P. Henneaux, P. Labeau, and J. Maun, "Blackout probabilistic risk assessment and thermal effects: impacts of changes in generation," *IEEE Trans. Power Systems*, vol. 28, pp. 4722–4731, Nov. 2013.
- [23] C. Asavathiratham, S. Roy, B. Lesieutre, and G. Verghese, "The influence model," *IEEE Control Systems Magazine*, vol. 21, pp. 52–64, Dec. 2001.
- [24] P. D. Hines, I. Dobson, E. Cotilla-Sanchez, and M. Eppstein, "'Dual Graph' and 'Random Chemistry' methods for cascading failure analysis," *46th Hawaii Intl. Conference on System Sciences*, HI, Jan. 2013.
- [25] C. Perrow, *Normal accident: living with high-risk technologies*, Princeton University Press, Princeton, 1999.
- [26] M. Begovic, D. Novosel, D. Karlsson, C. Henville, and G. Michel, "Wide-Area Protection and Emergency Control," *Proceedings of the IEEE*, vol. 93, pp. 876–891, May 2005.
- [27] V. Terzija, G. Valverde, D. Cai, P. Regulski, V. Madani, J. Fitch, S. Skok, M. Begovic, and A. Phadke, "Wide-area monitoring, protection, and Control of future electric power networks," *Proceedings of the IEEE*, vol. 99, pp. 80–93, Jan. 2011.
- [28] K. Sun, D. Zheng, and Q. Lu, "Splitting strategies for islanding operation of large-scale power systems using OBDD-based methods," *IEEE Trans. Power Systems*, vol. 18, pp. 912–923, May 2003.
- [29] K. Sun, K. Hur, and P. Zhang, "A new unified scheme for controlled power system separation using synchronized phasor measurements," *IEEE Trans. Power Systems*, vol. 26, pp. 1544–1554, Aug. 2011.
- [30] S. Lim, C. Liu, S. Lee, M. Choi, and S. Rim, "Blocking of zone 3 relays to prevent cascaded events," *IEEE Trans. Power Systems*, vol. 23, pp. 747–754, May 2008.
- [31] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, Jun. 1998.
- [32] R. Albert and A.-L. Barabási, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, Oct. 1999.
- [33] P. Hines, E. Cotilla-Sanchez, and S. Blumsack, "Do topological models provide good information about electricity infrastructure vulnerability?," *Chaos*, vol. 20, 033122, 2010.

Junjian Qi (S'12–M'13) received the B.E. and Ph.D. degree both in Electrical Engineering from Shandong University in 2008 and Tsinghua University in 2013. He visited Prof. Ian Dobson's group at Iowa State University in Feb.–Aug. 2012 and is currently a research associate at Department of EECS, University of Tennessee in Knoxville.

His research interests include blackouts, cascading failure, and state estimation.

Kai Sun (M'06–M'13) received the B.S. degree in automation in 1999 and the Ph.D. degree in control science and engineering in 2004 both from Tsinghua University, Beijing, China. He was a postdoctoral research associate at Arizona State University, Tempe, from 2005 to 2007, and was a project manager in grid operations and planning areas at EPRI, Palo Alto, CA from 2007 to 2012.

He is currently an assistant professor at the Department of EECS, University of Tennessee in Knoxville.

Shengwei Mei (SM'05) received the B.S. degree in mathematics from Xinjiang University, the M.S. degree in operations research from Tsinghua University, and the Ph.D degree in automatic control from the Chinese Academy of Sciences, in 1984, 1989, and 1996, respectively.

He is currently a professor at Tsinghua University. His research interests include power system analysis and control.