2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C)

Cascading Failure Path Prediction based on Association Rules in Cyber-Physical Active Distribution Networks

Chong Wang, Yunwei Dong, Pengpeng Sun, Yin Lu
School of Computer Science and Engineering,
Northwestern Polytechnical University,
Xi'an, PR China, 710072
wangchonw@mail.nwpu.edu.cn, yunweidong@nwpu.edu.cn,
jiea489756@163.com, Luyin@nwpu.edu.cn

***Abstract*—Cascading failures may lead to large scale outages, which brings about significant economic losses and serious social impacts. It is very important to predict cross-domain cascading failures paths for identification of weak nodes, which contributes to the control policies for preventing cascading failures and blocking their propagation between cyber domain and physical domain in cyber-physical active distribution networks. This paper proposes an algorithm based on the Frequent-Patterns-Growth (FP-Growth) to predict cascading failure paths, which predicts the potential failure node set by analyzing a large number of simulation datum and mining the hidden association relationship among datum. To demonstrate the effectiveness of the proposed cascading failure path prediction approach, an empirical study on a cyber-physical active distribution network, named CEPRI-CPS from Electric Power Research Institute of China, is performed, and the result shows the robustness of cyber-physical active distribution networks can be improved with prediction approach in this paper.**

***Keywords*—cascading failure path prediction, association rules, cyber–physical active distribution networks, FP-Growth algorithm**

I. Introduction

In recent years, a large number of distributed generators are encouraged to connect to distribution networks, namely active distribution networks (ADN) in many references [1][2]. As shown in Fig. 1, an ADN contains two interdependent networks, communication network (CN) and physical network (PN). The CN is composed of sensors, communication lines, data storage, computers, controllers, actuators, and other communication equipments. The PN is composed of generators, transmission lines, transformer substations, loads and other electrical equipments. All devices are connected via a certain topological structure. Some of the nodes in PN are responsible for providing power for nodes in CN, and some of the nodes in CN are responsible for collecting real-time information from nodes in PN and making controlling decisions for nodes in PN in order to keep PN in a stable state. For example, sensor nodes in CN can collect the information, such as current and voltage, from nodes of PN. Correspondingly, actuator nodes in CN affect the state of nodes in PN according to the collected information above. Therefore, ADN is a typical cyber-physical system called cyber-physical active distribution networks.

Since the CN and PN of ADN are interdependent, it is beneficial to the operation of ADN. However, the interdependency relationship can also bring some side effects. The failure of one node may not only be spread out in its own network, but also affect the interdependent network. This process continues until ADN reaches the steady state, which may cause cross-domain cascading failures (CCF), and then lead to large scale outages in ADN. There already existed many large-scale CCF occurrences worldwide. For example, CCF took place in Brazil in 2011, in India in 2012 and in Ukraine in 2015 [3]. Failures or abnormal operation of nodes in ADN will result in abnormal distribution of power flow and potential large-scale outages which cause significant economic losses and serious social impact [4]. Therefore, it is essential to analyze and predict the propagation of CCF to prevent the occurrence of CCF in ADN.

The propagation process of CCF is illustrated in Fig. 2. As shown in Fig. 2(a), when one node in CN fails, all the edges connected with it will also fail. This failure may propagate to the rest of nodes in CN through the connected relation and cause the failures of these nodes in CN, as shown in Fig. 2(b). Correspondingly, the node failures in CN may lead to the failures of nodes in PN, as shown in Fig. 2(c). After several propagation periods, ADN reaches a steady state, as shown in Fig. 2(d).

There are many factors which can cause the occurrence of CCF such as the close interdependence between PN and CN in AND, the improper operation caused by human or weather, etc. Due to the diverse affecting factors and complicated network structures, it is difficult to predict the CCF path reasonably using the traditional reliability-based method.

The close interdependence between CN and PN although improves the autonomy of ADN, it also increases the risk of the occurrence of CCF. The traditional methods in power system often analyze cascading failures using the circuit theory or the complex network theory [5] . These methods only focus on the failure analysis in PN and CN respectively, which do not consider the risk of interdependence between CN and PN [6]. In addition, it is difficult to calculate the power flow in ADN due to the complex topological structure and the access of distributed generators. The analysis methods, which are used for prediction of cascading failures in traditional distribution network, are not applicable to the analysis in ADN. Thus, it is necessary to adapt the traditional methods or find a new method to predict the cascading failure paths in ADN.

There must be causal relationship in the trajectory data because of the chain reaction [7] which is one of the salient features of cascading failures. This paper will consider PN and CN as a whole network and predicts all possible paths of CCF

978-1-7281-8915-4/20/$31.00 ©2020 IEEE
DOI 10.1109/QRS-C51114.2020.00083

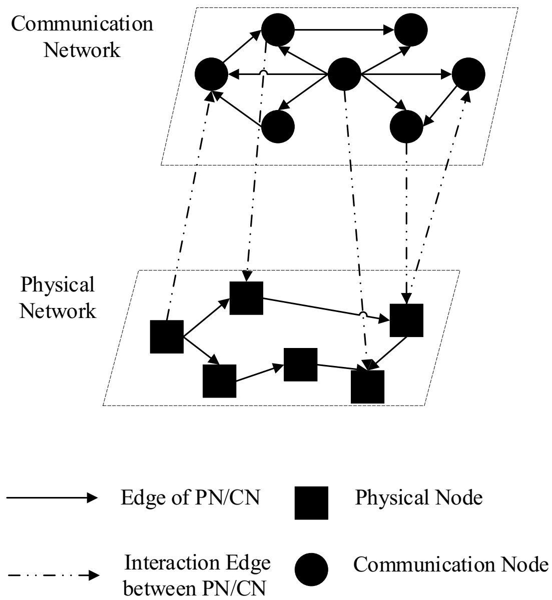

Fig. 1. Brief topological structure of ADN

in ADN by analyzing the association rule in the propagation process of CCF. FP-Growth is a data mining algorithm for association rule analysis. The association relationship hidden in data can be depicted by the association rules [8]. Furthermore, if there is an association relationship between two events, then it is possible to predict the occurrence of another event after one event occurs.

In this paper, we propose a method to predict all possible paths of CCF based on the association rules which are mined by the FP-Growth algorithm. Specifically, the hidden association rules are mined in the historical CCF data, and the association relationship among failure nodes in ADN are displayed by the association rules. Thus, our proposed prediction approach not only considers the failure propagation in PN and CN respectively, but also considers the interdependence between PN and CN in ADN.

The rest of this paper is organized as follows. Section II introduces the background of association rules and FP-Growth algorithm. Section III presents the algorithm for mining CCF paths. Section IV discusses the case study and simulation results. Section V describes the related work and Section VI concludes this paper.

II. BACKGROUND

Data Mining is an advanced procedure of data processing to extract potential, effective and understandable patterns from massive data with specific business objectives [9]. According to the different mining results, data mining can be divided into seven categories: feature rules, differentiation rules, classification, association rules, cluster, prediction, variation and deviation analysis. This paper focuses on the association rules mining, which aims to mine the association rules among a group of nodes for describing the closeness between data. Given a set of items $L = \{i_1, i_2, \cdots, i_m\}$, and a set of transactions $D = \{T_1, T_2, \cdots, T_n\}$ where transaction $T_i \subseteq L$ and each T_i has a unique identity TID. Let x be one item in set L, that the transaction T_i support x if T_i handles item x. And let Y be a set of items in L. it is said that the transaction T_i support Y if all the items in Y are handled by T_i. Thus, the association rule is formally defined as follows.

Definition 1. The form of association rule is noted as $x \rightarrow Y$, where x is an item in item set L, Y is also a node item in L that not contains x, that is $Y \subseteq L$ and $x \notin Y$. The rule $x \rightarrow Y$ is supported in the set of transactions D with the support factor $0 \leq c \leq 1$, which means at least $(100 * c)\%$ of transactions D support item x and item set Y.

The *support* degree d between X and Y is formally defined as formula (1), which is used to assess the association degree between X and Y, and it means the probability that X and Y occur synchronously in transactions D. The greater the probability is, the stronger the association is. In the formula (1), $N_{X\&Y}$ is the number of transactions in D that support X and Y synchronously, and N represents the number of all transactions in D.

$$d = \frac{N_{X\&Y}}{N} \tag{1}$$

The FP-Growth algorithm is an association rule mining algorithm, which can effectively discover the frequent item sets in data and the hidden association rules among historical data [10]. As demonstrated in [11], The FP-Growth algorithm performs better than other association rule mining algorithms

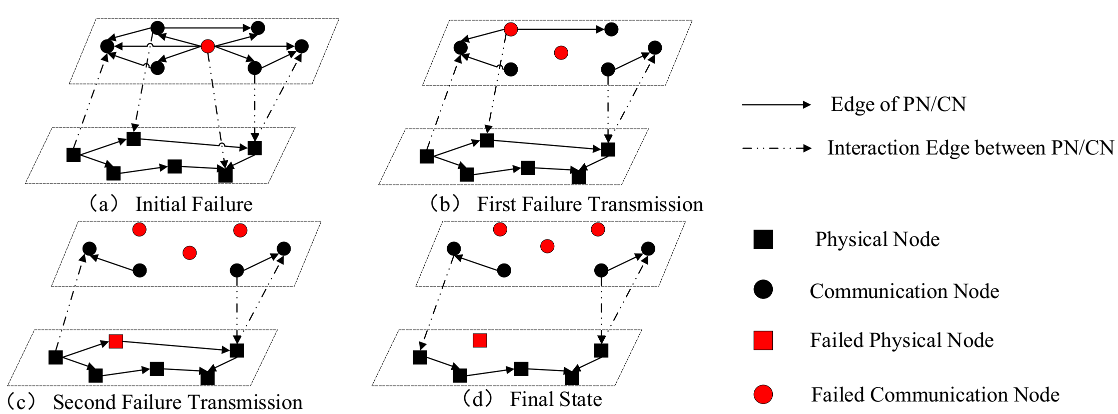

Fig. 2. Simple running example of CCF in an ADN

in time complexity. So the FP-Growth algorithm is chosen in this paper. The process of FP-Growth algorithm is mainly divided into two parts: FP-tree construction and recursive FP-tree mining. FP-tree is a kind of data structure, which compactly stores the transactions in the original data whose *support* degree are greater than the threshold values after scanning the data twice. FP-tree is similar to the prefix-tree, and the paths with the same prefix can be shared so as to achieve the purpose of data compression. Then, FP-tree is used to find out the conditional patterns of individual item. The conditional modes of each item are constructed into a new FP-tree, and frequent item sets are recursively mined from the new FP-tree.

A. FP-tree construction

The construction of the TP-tree in this paper is divided into three parts: a item set L, a root marked as null and tree node. Each node in FP-tree has three elements: item_name, count and node_link, $TN_i = \{\text{Item\_link}, \text{count}, \text{node\_link}\}$. Item_name means the item that the tree node represents, count means the number of transactions represented by the path from root to this tree node in the FP-tree and node_link points to the next tree node with the same Item_name in the FP-tree. If there is no same Item_name in the FP-tree, the node_link is set to null.

In order to build the FP-tree. We need to scan all the transactions in transactions set D that are needed to be mined in order to get number set $N = \{N_1, N_2, \cdots, N_m\}$, in which N_i is the number of transactions that support item i_i in L. After the number set N is got, all the item in item set L is sorted from big to small according to the values N_i in N. Then we scan all the transactions in D and we need to sort the element in each transaction T_i according to the order of node in the sorted node set L. In this step, we ignore the item whose number N_i is lower than threshold of *support* degree.

After each transaction in D is sorted, each element in transaction is inserted into FP-tree according to the sorted order. That means all the nodes that one transaction T_i handles are made up one path of the FP-tree. Node that, if FP-tree exists path that is identical or partially identical to the new path, these two paths' same parts need to be merged and the count of the identical node plus one.

B. User FP-tree to mine frequent items

The FP-tree is used to mine the frequent items after being constructed. The process of mining frequent items need to be applied to each item in item set L.

When we mine the frequent items of item i_n, all the tree nodes whose item_name is i_n should be got at first. Then we find the prefix path of these tree nodes. Prefix path of one tree node is that the tree nodes set from root to this tree node. It is assumed that prefix path $PT_i = \{TN_1, TN_2, \cdots, TN_m\}$. Then the prefix paths set of the tree nodes whose item_name is i_n can be described as $PTS = \{PT_1, PT_2, \cdots, PT_k\}$. The conditional tree can be built using prefix paths set PTS, whose step is the same as FP-tree construction. After constructing the conditional tree, perform recursive mining on the conditional tree. All the combinations of tree nodes in the conditional tree paths are the condition frequent sets.

III. CCF PATHS PREDICTION IN ADN

CCF paths can be viewed as a combination of failure nodes with different timestamps, so they are discrete time series. A failure node occurring at the initial time t_1 is called initial failure node S_1^1. S_1^1 is a communication node or a physical node. After a certain period of time, the failure is simultaneously passed to the nodes S_2^1 , $S_2^2, \cdots$, which are failure nodes at time stamp t_2. The failure keeps spreading out until the final failure nodes $S_n^1, S_n^2, \cdots$ at timestamp t_n, and ADN reaches the steady state.

The process of CCF path prediction algorithm divided into two steps: association rules mining and association rules matching. The process of association rules mining is shown in Fig. 3. Since raw data cannot meet the requirements of the algorithm, it should be preprocessed into a specific format. That is, the historical data needs to be classified according to the type of the initial failure node and extract the failed nodes at timestamp t_1, t_2. Then, FP-Growth algorithm is used to mine association rules in different kinds of preprocessed data.

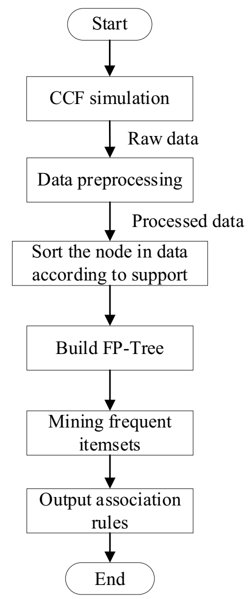

Fig. 3. Process of association rules mining

the Association rules matching is used to predict the path of CCF, the process of which is show in Fig. 4. In this process, the initial failure node is matched with them to obtain the node set that will fail at timestamp t_2. Next the failure nodes at timestamp t_2 are matched with the association rules to obtain the nodes that will be failed at timestamp t_3. Meanwhile, we ignore the node that has appeared in the previous node set. The process continues until no new failure nodes are matched. All the predicted failure nodes are regarded as the potential failure nodes in the propagation process of CCF, which form CCF paths.

A. Association rules mining

In this paper, the FP-Growth algorithm is applied to mine the corresponding associated nodes for each node in ADN for the first time. The function of associated nodes is to match the potential failure nodes in the process of predicting the CCF paths, and obtain the set of nodes that may fail in the next timestamp. Nodes that, in this process, each simulation of CCF should set the same initial failure node in one association rules mining. For example, if an ADN has 109 nodes, the association rules mining process should be applied for one hundred ninth times to each node.

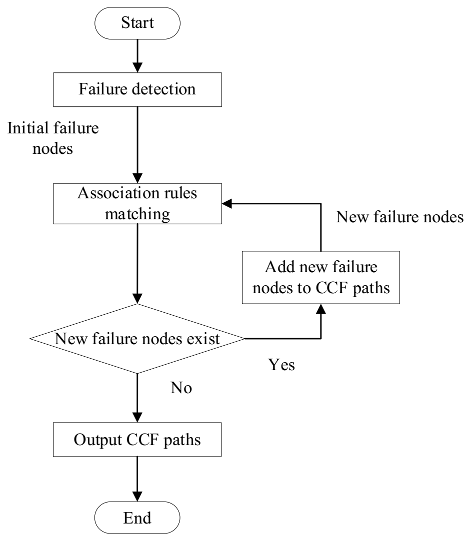

Fig. 4. Process of association rules matching

Definition 2. The item i_m in item set L means a node S_m in an ADN in the process of CCF path prediction algorithm. Set $RD_i = \{S_i^1, S_i^2, \cdots, S_i^n\}$ contains all the nodes that fail at the timestamp t_i, $RD_i \subseteq L$. RD_1 means the initial failure node which can have only one node, that is $RD_1 = \{S_1^1\}$. $CP_i = (RD_1, RD_2, \cdots, RD_n)$ is the node sequence which means the sets of failure nodes ordered by timestamps and can represent one CCF path when CCF happens at ith time. The set of CCF paths is defined as $CPL = \{CP_1, CP_2, \cdots, CP_i\}$.

Definition 3. The preprocess data $PD_i = \{RD_1, RD_2\}$, which means the preprocessing process is extracting the failure nodes that fails at timestamp t_1 and t_2 when CCF happens at ith time. PD_i can be viewed as a transaction T_i. The set of preprocess data is $PDS = \{PD_1, PD_2, \cdots, PD_i\}$. PDS can be viewed as a transaction set D.

Definition 4. The frequent items mean the number of transactions that support these items is more than the threshold of *support* degree. The frequent items can be described as $FI = \{S_{\mathrm{i}}, S_{\mathrm{m}}, \cdots, S_k\}$, $FI \subseteq L$, in which the first element if FI, S_{i} can be viewed as the item x, the rest elements if FI can be viewed as item set Y, the association rule is $S_{\mathrm{i}} \rightarrow \{S_{\mathrm{m}}, \cdots, S_k\}$. And *Frequentitemset* is the set of frequent items.

Definition 5. The association nodes set of each node is $AN_i = \{S_i, S_k, S_{k+1}, \cdots, S_m\}$. Each AN_i is the max length of frequent items mined by FP-Growth. Therefore, all the nodes set that are associated with each node in ADN can be describe as $ANS = \{AN_i, AN_2, AN_3, \cdots, AN_n\}$.

As shown in Algorithm 1, the process of association rules mining is present. The data preprocessing process is extracting the failure nodes that fails at timestamp t_1 and t_2. And then the preprocessed data is taken as input of FP-Growth. Many frequent items are obtained through FP-Growth Algorithm. Among them, The frequents item set which contains the initial nodes and the max number of items is considered as the associated nodes set which has an association relationship with the initial nodes.

Algorithm 1: Mining association rules

```
Input: CPL (the set of CCF paths with initial failure node S_1^1)
PD←null, PDS←null
for i←1 to length[CPL] do
  // no nodes fail at timestamp t2
  if CPL[i][1] == null then
     PD[i] = CPL[i][0]
  Else
     // preprocessed data is set to the nodes that fail at timestamp t1 and t2
     PD[i] = CPL[i][0] ∪ CPL[i][1]
  end if
PDS = PDS∪PD
end for
// take the preprocessed data as FP-Growth input
Frequentitemset = FP-Growth(PDS)
// take the max length of Frequent item as AN_i
max←0
for i←1 to length[Frequentitemset] do
  if Frequentitemset [j] contains initial failure node do
     if length[Frequentitemset [j]] > max do
        AN_i = Frequentitemset [j]
     end if
  end if
end for

Output: AN_i (Association nodes set)
```

The preprocessed data *PD* is shown in Table I. In Table I, there are five preprocessed data sets which construct the transaction set $PD = \{PD_1, PD_2, PD_3, PD_4, PD_5\}$. The initial failure node is set as S_1^1. Then the failure begins to spread out and five possible cases are obtained. The preprocessed data contains the failure nodes that fail at the timestamp t_1 and t_2. The data set PD is served as the input of FP-Growth Algorithm.

TABLE I. CASE OF CCF PREPROCESSED DATA

Timestamp / Transaction	t_1	t_2
PD_1	S_1^1	$S_2^1\ S_2^2\ S_2^3\ S_2^4$
PD_2	S_1^1	$S_2^2\ S_2^3\ S_2^4$
PD_3	S_1^1	$S_2^2\ S_2^3\ S_2^5$
PD_4	S_1^1	$S_2^1\ S_2^2\ S_2^3\ S_2^6$
PD_5	S_1^1	$S_2^3\ S_2^5$

If the threshold of *support* degree is set to c， for example c=70%, the association nodes mined by FP-Growth algorithm is $\{S_1^1, S_2^2, S_2^3\}$. So the association rule is $S_1^1 \rightarrow \{S_2^2, S_2^3\}$. The *support* degree between $\{S_1^1\}$and $\{S_2^2, S_2^3\}$ in the preprocessed data is $\frac{|\{PD_1, PD_2, PD_3, PD_4\}|}{|PD|} = 80\%$. It can be mined different *AN* which is dependent on thresholds. If the threshold of *support* degree is set to c=60%, the association nodes mined by FP-Growth algorithm is $\{S_1^1, S_2^1, S_2^2, S_2^3\}$. So the association rule is $S_1^1 \rightarrow \{S_2^1, S_2^2, S_2^3\}$. The *support* degree between $\{S_1^1\}$ and $\{S_2^1, S_2^2, S_2^3\}$ in the preprocessed data is $\frac{|\{PD_1, PD_2, PD_4\}|}{|PD|} = 60\%$, which is equal to the threshold of *support* degree. Hence, the threshold value of *support* degree determines the precision of FP-Growth algorithm. If the *support* degree threshold is set too high, the number of associated nodes is too small and some key nodes may be ignored. If the *support* degree threshold is set too low, there will be too many associated nodes, which leads to too many predicted paths. Therefore, it is necessary to select a proper threshold value which helps to predict CCF paths reasonably. The appropriate threshold can be obtained through expert knowledge or a large number of experiments.

B. Association rules matching

Once the association rule mining process is applied to each node in an ADN and get the associated nodes set of each node. the CCF path prediction will be conducted by matching the failure nodes with association rules. When the initial failure occurs, the initial failure node S_1^1 is matched with the association rules to obtain the potential failure node set $RD_2 = \{S_2^1, S_2^2, \cdots, S_2^n\}$. Then, the nodes in set RD_2 are matched with the association rules respectively to obtain the potential failure node set RD_3. Note that the newly failure nodes appeared in previous failure node sets should be removed in RD_3, in which only nodes never tried before should be kept. Continue matching failure node with association rules until no new potential failure nodes are found. All the sets of predicted failure nodes form a sequence of CCF paths $CP = (RD_1, RD_2, RD_3, ..., RD_n,)$.

The process of association rule matching algorithm is shown in Algorithm 2.

Algorithm 2: Matching association rules

```
Input：IFN (Initial failure node)
       ANS (Association nodes set of each node)
next_nodes = IFN, CP = next_nodes
//If new failure nodes are not matched, end the circulation
While (next _nodes EXISTS)
  next_nodes ← null
  for i←1 to length [ANS] do
    // Match IFN with ANS
    if IFN == ANS[i][0] then
      for j←1 to length [ANS[i]]
        //ignore the failure nodes appeared before
        if CP not contains (ANS[i][j]) then
          CP = CP ∪ ANS[i][j]
          next_nodes = next_nodes ∪ ANS [i][j]
        end if
      end for
    end if
  end for
end while

Output：CP (CCF_Paths)
```

IV. Case Study

A. Experimental setup

The simulation example of standard ADN *CEPRI-CPS* derived from the actual system is used as the experimental environment. There are 109 nodes in ADN, including 77 physical nodes (abbr. D1--D77) in PN and 32 communication nodes (abbr. R78--R109) in CN. The topology of ADN is shown in Fig. 5.

In the PN, nodes D_1-D_{22} and D_{72}-D_{76} are subnets of an industrial zone, nodes D_{23}-D_{42} and D_{77} are subnets of a residential zone, and nodes D_{43}-D_{62} are subnets of a commercial zone. The three subnets are connected to each other through the contact switch, and regard each other as its standby. In the simulation of CCF experiment， nodes D_{66}-D_{71} are the power supply nodes, in which nodes D_{66}-D_{68} are the photovoltaic power supply, node D_{69} is the gas turbine, node D_{70} is the battery energy storage device, and node D_{71} is the doubly-fed motor. Nodes D_{63}-D_{65} are external power supply nodes. Nodes D_1-D_4, D_{17}, D_{23}-D_{25}, D_{43}, D_{44}, D_{56}, D_{60}, D_{72}-D_{77} are distribution nodes and the rest are load nodes.

In the CN, there are 4 router nodes R_{80}, R_{81}, R_{82}, R_{83}, 5 switch nodes R_{79}, R_{84}, R_{85}, R_{86}, R_{87}, 1 server node R_{78}, several terminal nodes and communication lines. Three router nodes R_{80}, R_{81} and R_{82} are connected to four switch nodes R_{84}, R_{85}, R_{86} and R_{87} respectively. And the other router node R_{83} and switch node R_{79} are used to collect information of the three sub-stations and communicate with the control side. Nodes R_{84}-R_{109} are terminal nodes to collect data from nodes D_5-D_{25}, D_{36}, D_{40}, D_{43}, D_{45}, D_{49}, D_{57} and D_{72} in the PN.

B. Experimental results and analysis

To predict the CCF path, the CCF model [12] is used to simulate the propagation process of CCF in ADN. Each node in ADN is simulated with 1000 times for CCF. Note that the

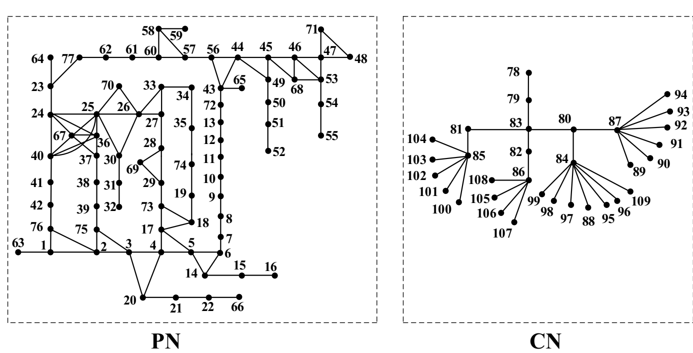

Fig. 5. Topological structure of ADN CEPRI-CPS

association rule will be $\{S_i\}\rightarrow\{\}$ when node S_i does not cause any cascading failures. This means that there is no associated nodes matched with S_i during the simulation process. Table II is an example of CCF raw data. The initial failure node is R_{86}. According to II, we can observe that CCF paths are the sets of failure nodes ordered by timestamps.

Before beginning prediction, it is necessary to preprocess the CCF raw data to mine the association rules for individual initial failure node. For example, take failure nodes at timestamp t_1 and t_2 into a CCF path PD_1= $\{R_{86}, D_{58}, D_{59}, D_{51}, D_{56}, D_{73}, D_{57}, D_{49}, D_{43}, D_{45}\}$ in which R_{86} is the initial failure node.

TABLE II. CASE OF CCF RAW DATA

Time Stamp	Failure Nodes
t_1	R_{86}
t_2	D_{58} D_{59} D_{51} D_{56} D_{73} D_{57} D_{49} D_{43} D_{45}
t_3	D_{50} D_{52} D_{65} D_{44}
t_4	R_{108} R_{106} R_{107} R_{105} R_{104} R_{82}
t_5	D_{42} D_{41} D_{40}

After all the raw data are preprocessed, each of 109 nodes will be mined for association rules in the ADN. Then, the association rules matched with each node are obtained. During the mining process, the *support degree* threshold is set to 50%. Table III introduces part of association rules of some nodes for the first initial node D_4. For example, D_4 is associated with node set $\{R_{94}, R_{92}, R_{90}, R_{91}, R_{93}, R_{89}, R_{88}, R_{87}\}$ according to the association rules shown in Table III. This means that after node D_4 fails, the nodes in the set $\{R_{94}, R_{92}, R_{90}, R_{91}, R_{93}, R_{89}, R_{88}, R_{87}\}$ are likely to fail.

TABLE III. CASE OF ASSOCIATION RULES

Initial Nodes	Associated Node Sets
D_4	R_{94} R_{92} R_{90} R_{91} R_{93} R_{89} R_{88}
R_{94}	D_{11} D_{60} D_{61} D_{62}
R_{92}	D_{50} D_{48} D_{46} D_{47} D_9
R_{90}	D_{26} D_{29} D_{28} D_{27} D_7
R_{91}	D_{37} D_{38} D_{39} D_8
R_{93}	D_{10} D_{52} D_{53} D_{54} D_{55}
R_{89}	D_{17} D_{18} D_{19} D_{20} D_6
R_{88}	D_5 D_2 D_3 D_1 D_4

During the prediction process, the initial failure node is set to D_4. According to the association rules, the potential failure nodes set at timestamp t_2 is $RD_2 = \{R_{94}, R_{92}, R_{90}, R_{91}, R_{93}, R_{89}, R_{88}, R_{87}\}$. The prediction algorithm continues to match each node in RD_2 with the association rules. Then, all failure nodes at timestamp t_3 that have an association relationship with the nodes in RD_2 are found. The associated failure nodes for each node in RD_2 are grouped into the node set RD_3. Table IV shows part of predicted sequence of CCF paths. Since the number of potential failure nodes is enormous in a large-scale ADN, we only list the predicted CCF data at timestamp t_1-t_3. According to predicted CCF paths, it can be concluded that when the node D_4 fails, it is highly likely to cause the failure of many terminal nodes in CN. Because of the significant failure propagation between nodes in both CN and PN, a large-scale CCF may occur in the ADN with 109 nodes.

From the perspective of ADN topological structure, when node D_4 fails, it is very likely to cause failures of some terminal nodes in CN that the D_4 supplies power to. This will result in the large-scale CCF because of the complex connection between massive nodes in ADN. In order to improve the robustness of the ADN and prevent the occurrence of CCF, some standby physical nodes in PN may be used to supply power to communication nodes in CN. Therefore, it is effective and feasible for topological structure optimization and operation maintenance of ADN using our proposed CCF path prediction algorithm.

TABLE IV. PREDICTION OF CCF PATH

Time Stamp	Failure Nodes
t_1	D_4
t_2	R_{94} R_{92} R_{90} R_{91} R_{93} R_{89} R_{88}
t_3	D_{11} D_{60} D_{61} D_{62} D_{50} D_{48} D_{46} D_{47} D_9 D_{26} D_{29} D_{28} D_{27} D_7 D_{37} D_{38} D_{39} D_8 D_{10} D_{52} D_{53} D_{54} D_{55} D_{17} D_{18} D_{19} D_{20} D_6 D_5 D_2 D_3 D_1 D_4

V. RELATED WORK

Many Research works have been done on the cascading failures in power system. The most extensive one is to find out the source of cascading failures from the perspective of network topology. The cascading failures can be divided into three classes according to the cause of the failures, the characteristics of critical events and other factors [13]. Mei et al [14] propose an OPA model which can describe the long-term slow dynamic behavior of the power system and analyze its cascading failures. Dobson et al [5] propose a CASCADE model which defines the probability distribution of normalized failure elements number. Zhou et al [15] provide a method for cascading failure analysis in power system based on fuzzy reasoning and fault tree theory, which has a good application prospect in the logic expression of cascading failures and the discrimination of development mode. Shen et al [16] combine the topological structure and power flow distribution of power system with Floyd algorithm to identify the shortest transmission path between the nodes of broken lines. The above work mainly focuses on the prediction or prevention of cascading failures for a single network. These studies can only improve the stability of either CN or PN rather than the whole connected ADN. Since ADN is a typical cyber-physical system, we must take the failure propagation, caused by the coupling relationship between CN and PN, into consideration.

Cao et al [17] establish the coupling model based on the DC power flow model, and studies the influence of routing strategy on the CCF of power system. Matsui et al [18] study the influence of inter-clustering coefficient on robustness of power system, which connects SCADA system with power networks following one-to-one connection relationship. Li et al [19] establish CCF model based on the complex system theory, and use the generation function theory and seepage theory to study CCF process. However, The research work above does not consider the impact of network dynamic topology on CCF in AND.

VI. Conclusion

The failure of a certain node in ADN may lead to large-scale CCF because of cascading failure propagation. This will have a great impact on the economic and social development. Therefore, the prediction of CCF paths could help engineer find a best controlling policy, and avoid cascade failures encounter among CN and PN, it will contribute to the secure operation of ADN. This paper proposes an algorithm based on the FP-Growth to predict cascading failure paths. This algorithm predicts the potential failure node set by analyzing a large number of experimental or simulation datum from actives power networks, and it could mines some hidden association relationship among datum which is described as association rules. When the active power networks takes counter a failures, and the failure is to be detected, and matches its features with association rules and finds its propagation paths as soon as possible to avoid cascading failures appear among cyber-Physical Active Distribution Networks. To demonstrate the effectiveness and feasibility of the path prediction algorithm, a numerical simulation against the ADN with 109 nodes was conducted. The result shows when some nodes fail, it is likely to cause large-scale power outages. More importantly, the topological structure of ADN can be optimized according to the prediction results. Thus, our proposed approach is of great help to the reliability of power supply and the prevention of CCF in ADN.

Acknowledgment

This work was supported by the National Key Research and Development Program of China (Basic Research Class)—Basic Theories and Methods of Analysis and Control of the Cyber–Physical Systems for Power Grid under Grant No. 2017YFB0903000.

References

[1] Bollen M H J, Sannino A. Voltage Control With Inverter-Based Distributed Generation[J]. IEEE Transactions on Power Delivery, 2005, 20(1):p.519-520.

[2] Pradit Fuangfoo, Wei-jen Lee, Ming-tse Kuo. Impact Study on Intentional Islanding of Distributed Generation Connected to Radial Subtransmission System in Thailand's Electric Power System[C]// Industry Applications Conference. IEEE, 2006.

[3] Cai Y , Cao Y , Li Y , et al. Cascading Failure Analysis Considering Interaction Between Power Grids and Communication Networks[J]. Smart Grid IEEE Transactions on, 2016, 7(1):530-538.

[4] Wang X , Zhu G , He R , et al. Survey of Cascading Failures in Cyber Physical Power System Based on Complex Network Theory[J]. Power System Technology, 2017.

[5] Dobson I , Chen J , Thorp J S , et al. Examining criticality of blackouts in power system models with cascading events[C]// System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on. IEEE, 2002.

[6] Schneider C M , André A. Moreira, José S. Andrade, et al. Mitigation of malicious attacks on networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108(10):3838-3841.

[7] Xukai Z , Wenying L , Yihan Y , et al. Evolution Mechanism and Preventing Strategies for Cascading Failure[J]. automation of electric power systems, 2008.

[8] Ullman J D . A Survey of Association-Rule Mining[C]// Discovery Science, Third International Conference, Ds, Kyoto, Japan, December. Springer-Verlag, 2000.

[9] Xiao Z, Shaoping M, Bo Z, et al. Data Minning: A Survey [J]. Pattern Recognition and Artificial Intelligence, 2001(01):50-57.

[10] Han J, Pei J,Yin Y.Mining frequent patterns without candidate generation[J]. ACM SIGMODRecord, 2000, 29(1):1-12.

[11] J. Han, J. Pei, Y. Yin, and R. Mao, "Mining frequent patterns without candidate generation: a frequent-pattern tree approach," Data Min. Knowl. Discovery, vol. 8, no. 1, pp. 53–87, Jan. 2004.

[12] Sturaro A , Silvestri S , Conti M , et al. Towards a Realistic Model for Failure Propagation in Interdependent Networks[J]. 2015.

[13] You-Bo L, Wei-Ting X U , Li-Jie D, et al. Power system cascading failure analysis theories and application II——key features of real cascading failures and revelation aspects[J]. power system protection and control, 2013, 41(10):146-155.

[14] Shengwei M , Fei H E, Xuemin Z, et al. An Improved OPA Model and the Evaluation of Blackout Risk[J]. Automation of Electric Power Systems, 2008.

[15] Zongfa Z, Xin A, Huiqiiong D, et al. A Method to Analyze Power System Cascading Failure Based on Fault Tree and Fuzzy Reasoning[J]. Power System Technology, 2006.

[16] Shen R , Liu D , Zhao J , et al. Weighted network model based recognition of dangerous lines under power flow transferring[J]. power system technology, 2012.

[17] Cao Y, Zhang Y, Bao Z. Analysis of cascading failures under interactions between power grid and communication network[J]. Dianli Zidonghua Sheb / Electric Power Automation Equipment, 2013, 33(1):7-11.

[18] Matsui Y, Kojima H, Tsuchiya T. Modeling the Interaction of Power Line and SCADA Networks[C]// IEEE International Symposium on High-assurance Systems Engineering. IEEE, 2014.

[19] Li W , Deng S, Li J, et al. Defense strategy of cascading failures between information network and physical power grid[J]. Gaodianya Jishu/high Voltage Engineering, 2013, 39(11):2714-2720.