Vulnerability Identification for Urban Power Networks Under Malicious Attacks Considering Topology Completion Uncertainties

Chengze Li, Wenxia Liu, Rui Cheng, and Yuze Yang

***Abstract*—Urban power network (UPN) topologies are partially observable due to invisible underground cables, leading to incomplete information under malicious attacks. To perform vulnerability identification for UPN under such attacks, it is of great importance to consider topology completion uncertainties in attack and defense strategies. To this end, this paper first proposes a UPN topology completion method, where a series of possible complete topologies associated with their corresponding probabilities are determined by means of Monte Carlo simulation and a knowledge-assisted hierarchical network partitioning algorithm, i.e., the improved SHRINK (I-SHRINK) algorithm. Subsequently, a bi-level stochastic optimization model is established to determine attack strategies by considering topology completion uncertainties, further identifying UPN vulnerability. The upper-level problem aims to maximize the expected load shedding risk, while the goal of lower-level problem is to minimize the total load shedding. Case studies are performed in a real UPN to verify the effectiveness and superiority of the UPN topology completion and vulnerability identification methods.**

***Index Terms*—Stochastic optimization, urban power network (UPN), vulnerability identification, topology completion, uncertainty, malicious attack.**

I. Introduction

In recent years, the low-probability but increasing malicious attacks have posed significant threats to urban power networks (UPNs) [1] - [3], causing high-impact damage and affecting their safe and reliable operation. The Wall Street Journal reported 274 incidents of malicious attacks on UPNs in the USA from 2012 to 2014 [4]. In 2019, Venezuela experienced a series of malicious attacks that lasted more than 24 hours [5], while threats to Iranian nuclear power plants and Ukrainian power grid have further highlighted UPN security concerns [6]. Modeling malicious attacks, quantifying their consequences, and identifying UPN vulnerabilities under such threats have garnered increasing attention in recent years.

To identify the UPN vulnerabilities under malicious attacks, existing studies focus on locating critical components through topological structure analysis [7]. References [8] - [10] identify vulnerabilities based on complex network indicators including degree, connectivity, and centrality of node. In [11], a sequential attack strategy is proposed to identify vulnerabilities in the power system using the degree of node and load classification. Reference [12] further analyzes the cascading failures triggered by vulnerable nodes.

The vulnerability identification based on the aforementioned topological structure does not fully capture the operational states and characteristics of UPNs, leading to inaccuracies in vulnerability identification. By taking into account the operational states and characteristics of UPNs, some research works [13], [14] formulate malicious attack problems as typical bi-level programming problems in attack-defense scenarios. These works simulate strategic interactions between the attacker and defender to identify UPN vulnerabilities by assessing the operational consequences of malicious attacks. In [15] - [17], the attack-defense optimization problems are formulated with the objective of maximizing system load and generator output losses with limited attack resources.

However, these works do not consider the incomplete UPN information when identifying vulnerabilities under malicious attacks. Data-driven approaches are proposed for identifying the parameters of UPN by using a large number of historical operation data, which is challenging and difficult to obtain in reality [18], [19]. References [20] and [21] indicate that attackers find it challenging to determine the exact operational status of the power system. However, they can consider potential statuses and devise effective attack strategies using statistical tests, graph theory, and boundary flow shift factors.

The aforementioned studies mainly deals with vulnerability identification under complete UPN topologies with incomplete parameter information, but does not consider the practical possibility of incomplete UPN topologies. Many UPN lines in highly developed cities are underground cables, making UPN topologies unobservable. In actual malicious attack scenarios, attackers often lack complete information about

Manuscript received: November 21, 2024; revised: March 21, 2025; accepted: May 30, 2025. Date of CrossCheck: May 30, 2025. Date of online publication: July 7, 2025.

This work was supported by the China Postdoctoral Science Foundation (No. 2024M750892) and the Fundamental Research Funds for the Central Universities (No. 2025JC004).

C. Li, W. Liu, R. Cheng (corresponding author), and Y. Yang are with the Department of Electrical & Computer Engineering, North China Electric Power University, Beijing, China (e-mail: upclcz007@163. com; liuwenxia001@163. com; ruicheng@ncepu.edu.cn; 1835176017@qq.com).

DOI: 10.35833/MPCE.2024.000803

these topologies. For example, the UK National Grid [22] has approximately 4500 miles of overhead lines and more than 900 miles of underground cables in cities, resulting in an underground cable proportion of around 20%. Thus, attackers can only reconstruct incomplete UPN topologies into complete configurations to formulate attack strategies and identify vulnerabilities.

References [23] and [24] propose an adaptive algorithm that employs a reweighting scheme to complete the graph-edge structure based on incomplete heterogeneous data from a collection of related topologies. In [25], an algorithm called SHRINK-H is introduced for topology partitioning in incomplete networks, which can infer potential hidden lines based on the partitioning results. Reference [26] employs a similar algorithm to perform topology partitioning on the airborne ad hoc network for observation. It combines community detection and inference results to construct a mixed-connection probability matrix, followed by topology correction on the estimated network.

All these works [23]-[26] are not related to UPN, and few studies have been conducted on completing UPN topologies. Meanwhile, existing studies in other systems often focus on identifying the most reasonable topology based on specific metrics. However, in practice, due to uncertain events such as cable route deviations caused by installation geography or limited road space, the UPN topologies in real world may not have been built according to the economically optimal planning scheme, and some suboptimal topologies may still be real solutions. The existence of multiple plausible UPN topologies introduces uncertainty for malicious attackers, making it impractical to rely on a single complete topology for attack strategies and vulnerability identification. By reconstructing multiple highly plausible topologies, this uncertainty can be significantly reduced, enhancing the effectiveness of attacks.

Uncertainty challenges have been widely studied in power systems, e.g., stochastic optimization [27], [28]. In [29], a series of uncertain scenarios are generated for the growth of electricity and hydrogen demands, establishing a multistage stochastic co-planning model for the integrated power distribution and hydrogen system. Reference [30] proposes the optimal power system defense planning to minimize the expected losses in different attack scenarios.

Motivated by the shortcomings of prior work, prior to the vulnerability identification for UPN under malicious attacks, it is essential to propose a feasible approach for completing the UPN topology to mitigate the associated uncertainties. In this paper, we propose a complete topology sampling model, enabling attackers to reconstruct high-rationality complete UPN topologies to develop attack strategies. Next, we propose a bi-level stochastic optimization model to identify UPN vulnerabilities. The main contributions are as follows.

1) A UPN topology completion method is proposed to take into account topology completion uncertainties by means of Monte Carlo simulation and a knowledge-assisted hierarchical network partitioning algorithm, i. e., the improved SHRINK (I-SHRINK) algorithm, thus constructing a set of reasonable complete topologies of UPN.

2) The I-SHRINK algorithm is proposed to ensure the high rationality of the extracted topology. Given that nodes with more similar properties tend to have tighter connections, the I-SHRINK algorithm leverages topology partitioning detection to uncover similarities among nodes, incorporating UPN factors such as node distances and visible signs of hidden cables. Based on the partitioning detection, we evaluate the complete cable rationality. In addition, the operational rationality and complete cable rationality are considered to enhance the UPN topology completion from a knowledge-based perspective.

3) A bi-level stochastic optimization model is developed to analyze attack strategies, identifying vulnerable nodes amid UPN topology uncertainties. The upper-level problem aims to maximize the expected load shedding risk to identify the optimal attack strategy, while the lower-level problem with defender emergency response strategy aims to minimize the total load shedding. By applying the Karush-Kuhn-Tucker (KKT) conditions, the original model is transformed into a single-layer mixed-integer linear program (S-MILP) one. In the case study, we rigorously examine the systematic differences between the true results and the vulnerability identification results obtained under the topology completion uncertainties considered in this study.

The remainder of this paper is organized as follows. Section II introduces the UPN topology completion method. Section III introduces the topology rationality assessment. Section IV develops the bi-level stochastic optimization model for vulnerability identification under malicious attacks, and Section V presents numerical case studies. At last, Section VI provides conclusions and points out directions for future research.

II. UPN Topology Completion Method

A. Incomplete UPN Modeling and Topology Completion

This section first models the incomplete UPN topology and then presents the guiding principles for topology completion from the attacker perspective. The UPN considered in this paper mainly includes the urban transmission network (UTN) and its terminal high-voltage urban distribution network (HV-UDN). Incomplete lines include UTN cables and transmission cables connecting the UTN and HV-UDN.

Let $\mathcal{G}\langle \mathcal{V}, \mathcal{L} \rangle$ represent the incomplete UPN topology, where $\mathcal{V}$ is the set of stations, and $\mathcal{L}$ is the set of visible overhead lines. Define $\mathcal{V}=\left(\mathcal{V}_G, \mathcal{V}_S\right)$, where $\mathcal{V}_G$ and $\mathcal{V}_S$ are the sets of power plants and substations, respectively. By applying certain completion principles, an attacker aims to change $\mathcal{G}\langle \mathcal{V}, \mathcal{L} \rangle$ into a complete topology $\mathcal{G}_T=\langle \mathcal{V}, \mathcal{L}, \mathcal{C}_T \rangle$ with high accuracy, where $\mathcal{C}_T$ is the set of complete cables. However, as shown in Fig. 1, the potential complete topology $\mathcal{G}_T$ is not unique. That is, there could be multiple plausible complete topologies, defined as $\mathcal{G}_{T1}, \mathcal{G}_{T2}, ..., \mathcal{G}_{Tn}$.

With respect to the incomplete UPN topology $\mathcal{G}$ observed by the attacker, we define its corresponding real complete topology observed by the defender as $\mathcal{G}_0=\langle \mathcal{V}, \mathcal{L}, \mathcal{C}_0 \rangle$, where $\mathcal{C}_0$ is the set of actual cables. Given that attackers are unaware

of $\mathcal{G}_0$, they can only take advantage of logical principles to derive a set $\mathcal{G}_T$ with high rationality. This enables them to strategically minimize the fault tolerance rate of resource allocation in various $\mathcal{G}_T$ scenarios under malicious attacks.

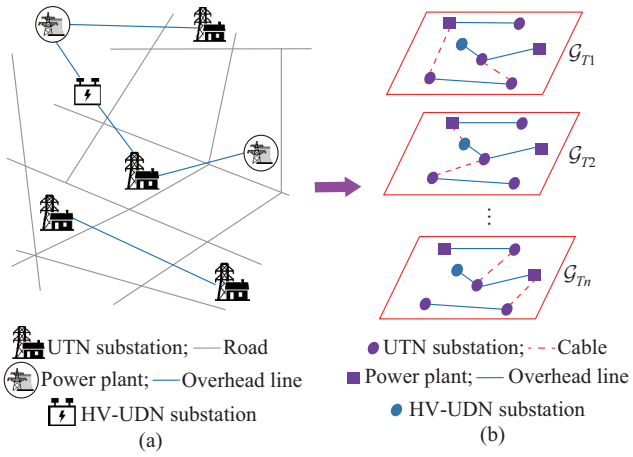


Fig. 1. An incomplete UPN with possible complete topologies. (a) Incomplete UPN. (b) Possible complete topologies.

Let R_T represent the topology rationality of $\mathcal{G}_T$ from the perspective of the attacker. $\mathcal{G}_T$ with a high R_T should meet the following conditions: it is in accordance with the UPN design specifications, ensures the economic efficiency of operation, and has a reasonable path for each cable in $\mathcal{C}_T$. The complete cable rationality of $\mathcal{C}_T$ is mainly judged by the presence of sufficient visible signs along the cable paths.

Note that in this paper, all dynamic parameters, which change with the topology $\mathcal{G}_T$, are denoted with a superscript T; while the static parameters, which are related to $\mathcal{G}_0$ or remain unchanged with respect to $\mathcal{G}_T$, do not have this superscript.

B. Complete Topology Sampling with I-SHRINK Algorithm

In this subsection, we use the Monte Carlo simulation to derive a set of complete topologies Ω_T with high rationality, and $|\Omega_T| = N_W$. For $\forall \mathcal{G}_T \in \Omega_T$, its rationality R_T satisfies $R_T > \xi$, where ξ is the pre-set threshold. Ω_T functions as a stochastic scenario ensemble to develop the attack strategies in Section IV, with the goal of maximizing the attack effectiveness while minimizing uncertainties associated with topology completion.

First, we use the Monte Carlo simulation to complete $\mathcal{G}$ into $\mathcal{G}_T$ by randomly extracting node connections. Then, the I-SHRINK algorithm is proposed to assess the topology rationality $\mathcal{R}_T$ of $\mathcal{G}_T$. If $\mathcal{R}_T$ is high, $\mathcal{G}_T$ will remain in Ω_T; otherwise, it will be extracted.

The specific steps are as follows.

Step 1: let X represent the set of node pairs in $\mathcal{G}$ that do not have overhead line connections. For $\forall (i,j) \in X$, the attacker randomly generates underground cable connections to form the complete cable set $\mathcal{C}_T$, thus completing $\mathcal{G}$ into $\mathcal{G}_T$.

Step 2: partition the incomplete UPN topology, and calculate the structural rationality, operational rationality, and complete cable rationality.

① Incomplete UPN topology partitioning

From the structural perspective, partition the incomplete UPN topology $\mathcal{G}$ using the I-SHRINK algorithm. Considering both the density of cable signs on the roads between nodes and the distance between nodes, we propose an evaluation method to assess the existence of concealed lines, where the endpoints of these lines may be located in the same partition or different partitions.

② Structural rationality calculation

Determine whether $\mathcal{G}_T$ is reasonable in terms of structural features. If it is reasonable, enter ③; otherwise, discard $\mathcal{G}_T$ and return to *Step 1* for re-extraction.

③ Operational rationality calculation

The operational simulation of $\mathcal{G}_T$ is conducted to determine the normal operational state of the system, followed by calculating the operational rationality I_T^1. If $I_T^1 > \lambda_1$, where λ_1 is a base value, enter ④; otherwise, discard $\mathcal{G}_T$ and return to *Step 1* for re-extraction.

④ Complete cable rationality calculation

Based on the partitioning result in ①, the rationality of the existence of each cable l_{ij} in the complete cable set $\mathcal{C}_T$ can be assessed. The complete cable rationality is denoted as I_T^2: if $I_T^2 > \lambda_2$, where λ_2 is a base value, enter *Step 3*; otherwise, discard $\mathcal{G}_T$ and return to *Step 1* for re-extraction.

Step 3: calculate the topology rationality of $\mathcal{G}_T$, i.e., $R_T = I_T^1 I_T^2$. If $R_T > \xi$, retain the current topology $\mathcal{G}_T$ and add it to Ω_T; otherwise, discard $\mathcal{G}_T$ and return to *Step 1* for re-extraction.

Step 4: determine whether the number of topologies in Ω_T reaches N_W. If yes, save and output the set Ω_T; otherwise, return to *Step 1*.

The flowchart detailing the above process can be found in Supplementary Material A. The details of *Step 2* are discussed in Section III.

III. Topology Rationality Assessment

A. UPN Topology Partitioning by I-SHRINK Algorithm

Reference [25] proposed an algorithm called SHRINK-H for clustering and partitioning incomplete-connection networks based on distance metrics and modularity Q [31], [32]. Furthermore, an algorithm using a greedy approach called SHRINK-G is introduced for higher computational efficiency, indicating that the tightly-connected nodes share similar properties and belong to the same partition [25]. Reference [33] later enhanced this with an algorithm called D-SHRINK, which can identify overlapping partitions and outliers.

In this subsection, we introduce an I-SHRINK algorithm to partition incomplete UPN topologies, facilitating the calculation of I_T^2. First, we consider the characteristics of power nodes, the distances between them, and the density of cable signs along the roads to propose a partitioning method for the incomplete UPN topology $\mathcal{G}$. Considering the partitioning results and structural characteristics of the UPN, along with the relationship between transmission distances and node voltage levels, we propose an evaluation index for hidden lines between nodes and calculate I_T^2.

The cosine similarity effectively denotes the local connectivity density between any two nodes in a weighted network [25]. Motivated by this concept, for $\forall i,j \in \mathcal{V}$, we define the modular correlation degree δ_{ij} between nodes i and j as:

$$\delta_{ij}=\frac{\sum_{x\in\Gamma(i)\cap\Gamma(j)}\omega(i,x)\omega(j,x)}{\sqrt{\sum_{x\in\Gamma(i)}\omega^2(i,x)}\sqrt{\sum_{x\in\Gamma(j)}\omega^2(j,x)}} \tag{1}$$

where $\omega(i,x)$ is the modular similarity between nodes i and x; and $\Gamma(i)$ is the set of nodes in $\mathcal{G}$ with high modular similarity to node i, and if $\omega(i,x)\geq\gamma$ (γ is the minimum correlation constant), $x\in\Gamma(i)$. The above explanation also applies to node j.

$\omega(i,x)$ can be expressed as:

$$\omega(i,x)=\tau(i,x)d(i,x) \tag{2}$$

where $\tau(i,x)$ and $d(i,x)$ are the presence of underground cable and distance similarity between nodes i and x, respectively

$d(i,x)$ is expressed as:

$$d(i,x)=1-\mathrm{e}^{-dis(i,x)} \tag{3}$$

where $dis(i,x)$ is the length of line if there is an overhead line between nodes i and x, and otherwise, $dis(i,x)$ is the average distance of road between nodes i and x.

While attackers cannot directly observe the underground cables between nodes i and x, cable signs on main roads between them indicate their presence. Therefore, $\tau(i,x)$ is determined by the density of these observable signs as:

$$\tau(i,x)=\begin{cases}1-\mathrm{e}^{-a_{ix}} & \lambda_{ix}=1\\ 1 & \lambda_{ix}=0\end{cases} \tag{4}$$

where λ_{ix} is a constant indicating whether there is at least one direct main road between nodes i and x: if yes, $\lambda_{ix}=1$, otherwise, $\lambda_{ix}=0$; and a_{ix} is the support parameter for the existence of a cable connection between nodes i and x.

Nodes i and x may be connected by more than one road, and a_{ix} can be calculated as:

$$a_{ix}=\frac{1}{Nl_{ix}}\sum_{l\in L_{ix}}a_l \tag{5}$$

where L_{ix} is the set of roads between nodes i and x; Nl_{ix} is the total number of roads in L_{ix}; and a_l is the support parameter indicating whether a cable is laid under road l between nodes i and x.

Cable signs differ in their significance for indicating the presence of cables. As shown in Table I, they are classified into three classes (classes A, B, and C) according to their importances from highest to lowest. The value of a_l can be calculated based on the numbers of cable signs for these three classes near road l:

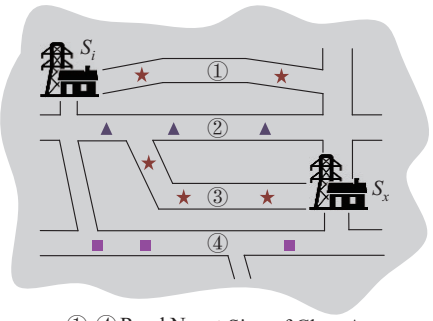
$$a_l=\mathbb{E}_l\left(\varepsilon_1 m_l^a+\varepsilon_2 m_l^b+\varepsilon_3 m_l^c\right) \tag{6}$$

where m_l^a, m_l^b, and m_l^c are the numbers of cable signs for classes A, B, and C near road l, respectively, and ε_1, ε_2, and ε_3 are their corresponding weights; and $\mathbb{E}_l$ is the overall expert score on the rationality of cable installation along road l: a higher $\mathbb{E}_l$ indicates lower difficulty in laying cables.

TABLE I
CLASSIFICATION OF OBSERVABLE CABLE SIGNS

Class of cable sign	Type of visible cable sign	Position
A	Signs for exposed cable section	Along the road
B (cable pipe network)	Signs for maintenance, vents, entrances and exits, alarms/reminders, etc.	Along the road or pavement
C (utility tunnel)	Signs for maintenance, vents, entrances and exits, alarms/reminders, etc.	Along the road or pavement

Use the example of Fig. 2 to better understand the presence of concealed cable between substations S_i and S_x.



①-④ Road No.; ★Sign of Class A
▲Sign of Class B; ■Sign of Class C

Fig. 2. Road network and cable signs between substations S_i and S_x.

Assume that $\varepsilon_1=0.6$, $\varepsilon_2=0.3$, and $\varepsilon_3=0.1$. Figure 2 illustrates four roads ①-④ between substations S_i and S_x, the values $\mathbb{E}_l$ of which are 0.75, 0.65, 0.77, and 0.35, respectively. Then, a_l can be calculated based on the values of m_l^a, m_l^b, and m_l^c. Finally, $a_{ix}=1.16$ and $\tau(i,x)=0.69$ can be calculated.

The distance similarity $d(i,x)$ is expressed as:

Referring back to (1), if $\mathcal{G}$ is partitioned as $\mathcal{G}=\{\mathcal{G}_1,\mathcal{G}_2,...,\mathcal{G}_k\}$, we define the distance-based modular measurement Q_d to judge the partitioning results [31], [32]:

$$Q_d=\sum_{s=1}^{k}Q_d^s=\frac{D_s^I}{D_T}-\left(\frac{D_s^G}{D_T}\right)^2 \tag{7}$$

where Q_d^s is the modular similarity of partition $\mathcal{G}_s$; $D_s^I=\sum_{i,j\in\mathcal{G}_s}\delta_{ij}$ is the sum of modular correlation degrees between any pair of nodes in $\mathcal{G}_s$; $D_s^G=\sum_{i\in\mathcal{G}_s,j\in\mathcal{V}}\delta_{ij}$ is the sum of modular correlation degrees between one node in $\mathcal{G}_s$ and the other node in $\mathcal{G}_0$; and $D_T=\sum_{i,j\in\mathcal{V}}\delta_{ij}$ is the sum of modular correlation degrees between any two nodes in $\mathcal{G}_0$.

If the two partitions $\mathcal{G}_s$ and $\mathcal{G}_t$ are combined into one partition, then the change of Q_d is:

$$\Delta Q_d=Q_d^{s\cup t}-Q_d^s-Q_d^t=\frac{2D_{st}}{D_T}-\frac{2D_s^I D_t^I}{D_T^2} \tag{8}$$

where $Q_d^{s\cup t}$ is the modular similarity of the partition formed by merging $\mathcal{G}_s$ and $\mathcal{G}_t$; and $D_{st}=\sum_{i\in\mathcal{G}_s,j\in\mathcal{G}_t}\delta_{ij}$.

Therefore, if $\mathcal{G}_1,\mathcal{G}_2,...,\mathcal{G}_k$ are combined, then ΔQ_d is:

$$\Delta Q_d = \frac{\sum_{s,t\in\{1,2,\ldots,k\},s\neq t} 2D_{st}}{D_T} - \frac{\sum_{s,t\in\{1,2,\ldots,k\},s\neq t} 2D_s^I D_t^I}{D_T^2} \tag{9}$$

Lastly, based on [32], we define ε-approximation microcommunity in the incomplete UPN topology $\mathcal{G}$ as follows.

Definition 1 ε-approximation microcommunity. For $\mathcal{G}=(\mathcal{V},\mathcal{L})$, $\mathcal{SG}_T(i)=(\mathcal{V}',\mathcal{L}')$ is a subgraph of $\mathcal{G}$. $\mathcal{SG}_T(i)$ is said to be an ε-approximation microcommunity of $\mathcal{G}$ if: ① $i\in\mathcal{V}'$; ② for $\forall j\in\mathcal{V}'$, $\exists i\in\mathcal{V}'\cap\left(j\leftrightarrow_{\varepsilon} i\right)$, where $\left(j\leftrightarrow_{\varepsilon} i\right)$ indicates that (i,j) is a tightly-coupled node pair; and ③ $\nexists j\in\mathcal{V}$, $\left\{j|i\in\mathcal{V}'\cap j\in\mathcal{V}\cap j\notin\mathcal{V}'\cap\left(j\leftrightarrow_{\varepsilon} i\right)\right\}=\varnothing$. The tightness criterion ε is expressed as:

$$\varepsilon=\max\left\{\sigma_{xy}|\left(x=i,y\in\Gamma(i)\backslash\{i\}\right)\cup\left(x=j,y\in\Gamma(j)\backslash\{j\}\right)\right\} \tag{10}$$

According to the potential similarity embodied by σ_{ij} between nodes i and j, $\mathcal{G}$ is partitioned based on the following steps.

Step 1: each generator node in the observable grid is treated as an initial partition to ensure the power supply rationality for each partition.

Step 2: for each node, identify and initialize its ε-approximation microcommunity.

Step 3: calculate the initial value ΔQ_d of each ε-approximation microcommunity: if $\Delta Q_d>0$, the microcommunity is regarded a generalized node and $\mathcal{G}_T$ shrinks in a generalized way.

Step 4: repeat *Steps 1-3* for the shrunk grid until each microcommunity satisfies $\Delta Q_d<0$.

Step 5: obtain the final partitions of $\mathcal{G}$.

The described steps achieve high efficiency by rapidly reducing the network scale, ensuring that each node is accessed precisely once in each iteration and that the final partitioning outcome is independent of the node access sequence. As illustrated in Fig. 3, with the incomplete UPN, the concept of partitioning becomes visually apparent. Partitioning incomplete UPN topology depends on the modular correlation degree between nodes, resulting from the interaction of multiple layers, as shown in red and blue blocks in Fig. 3. Nodes that are closer together in the same partition are likely to have more cable signs on the connecting roads, indicating a higher probability of hidden cable connections.

B. Complete Cable Rationality Calculation

Based on the partitioning results, this subsection calculates the complete cable rationality I_T^2 between nodes, corresponding to ④ in *Step 2* in Section II-B.

Due to the higher modularity Q_d and closer connectivity $\omega(i,x)$ between nodes in each partition of $\mathcal{G}$, a node is more likely to have concealed cables connected to other nodes in the same partition, and less likely to those outside this partition [34]. This is primarily attributed to the longer transmission distances and lower modular correlation degree with the nodes outside the partition. This subsection calculates the complete cable rationality I_T^2 between nodes, corresponding to ④ in *Step 2* in Section II-B.

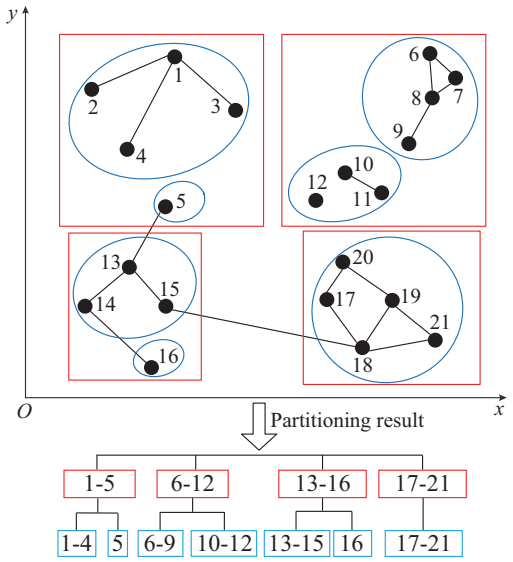


Fig. 3. Partitioning result of an incomplete UPN topology.

1) For nodes i and j located in the same partition

The justification for concealed connections between nodes i and j in the same partition is determined by $\omega(i,j)$ and the structural attributes of nodes i and j:

$$\gamma_{ij}=\omega(i,j)\left(\frac{k_i k_j}{\sum_{v\in lo_{ij}} k_v^2}\right)^{\alpha}\frac{1}{1+\left(\mu_i-\mu_j\right)^2} \tag{11}$$

where lo_{ij} is the community partition shared by nodes i and j; k_i, k_j, and k_v are the degrees of nodes i, j, and v, respectively; $\alpha\geq 0$ is the conversion parameter; and μ_i and μ_j ($\mu_i,\mu_j\in[0,1]$) are the conversion parameters based on the voltage levels of nodes i and j, respectively: if the voltage level of node i is higher, then $\mu_i>\mu_j$.

2) For nodes i and j located in different partitions

When nodes i and j are located in different partitions, a low value of $\omega(i,j)$ indicates a minimal presence of concealed cables and suggests close distances between them. This implies that a strong functional correlation is essential for the existence of concealed cable connections. For example, if both nodes i and j are high-voltage substations, or if one of them is a large power plant, there is likely a long-distance power transmission between them.

Consequently, the parameter γ_{ij} is expected to have a positive correlation with the voltage levels of nodes i and j, while exhibiting a negative correlation with the distance between them:

$$\gamma_{ij}=\omega(i,j)\left(\frac{k_i}{\sum_{v\in lo_i} k_v}\frac{k_j}{\sum_{r\in lo_j} k_r}\right)^{\beta}\mu_i\mu_j\frac{1}{2}\frac{\bar{d}_i+\bar{d}_j}{d(i,j)} \tag{12}$$

where lo_i and lo_j are the community partitions for nodes i and j, respectively; $\bar{d}_i$ and $\bar{d}_j$ are the average distances of observable overhead lines between nodes i, j and other nodes, respectively; and β is the conversion parameter, with $\beta\geq 0$.

Based on (11) and (12), the complete cable rationality I_T^2 is calculated as:

$$I_T^2=\frac{1}{NL_T}\sum_{l_{ij}\in\mathcal{C}_T}\gamma_{ij} \tag{13}$$

where NL_T is the number of cables in the completion cable set $\mathcal{C}_T$.

Compared with the traditional SHRINK algorithms, the I-SHRINK algorithm thoroughly considers the differentiated characteristics of power systems in terms of the incomplete UPN topology partitioning and evaluation process.

C. Structural Rationality Calculation

In the Ω_T generation process (corresponding to ② in *Step 2* in Section II-B), we emphasize that $\mathcal{G}_T$ complies with established design standards, such as IEEE Power Transmission and Distribution Standards [35]. To calculate the structural rationality of $\mathcal{G}_T$, two judgment factors are considered.

1) Connectivity

To leverage the efficiency, the breadth-first search (BFS) algorithm [36] is employed to identify the path between nodes i and j, $\forall i,j\in\mathcal{V}$. If a path exists for all pairs of nodes $\forall i,j\in\mathcal{V}$, indicating no isolated components in the completed topology, then the connectivity index $\eta_T^1=1$; otherwise, $\eta_T^1=0$:

$$\eta_T^1=\begin{cases}1 & \forall i,j\in\mathcal{V},r_{ij}^T=1\\ 0 & \exists i,j\in\mathcal{V},r_{ij}^T=0\end{cases} \tag{14}$$

where r_{ij}^T is an indicator of whether nodes i and j meet the connectivity in $\mathcal{G}_T$: if nodes i and j are connected, $r_{ij}^T=1$; otherwise, $r_{ij}^T=0$.

2) Power supply rationality

For node $\forall i\in\mathcal{V}_S$, define its voltage level as E_i. For $\mathcal{G}_T$, the power supply mode shown in Fig. 4 is not reasonable, where the lower-level substation S_2 serves as a power source for the higher-level substation S_3, effectively functioning as a boost station. In a transmission network, only the substation directly connected to the power plant can function as a boost station. This is crucial because the power plant generates electricity at a relatively low voltage, which needs to be stepped up to reduce power loss during long-distance transmission. Thus, the power supply mode involving S_3 and S_2 is not reasonable.

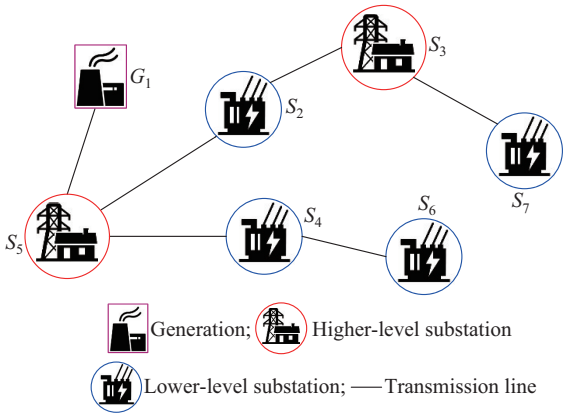


Fig. 4. Unreasonable power supply mode.

For $\forall i\in\mathcal{V}_S$ and $\forall j\in\mathcal{V}_G$, search for their connection paths and define the path set as Φ_{ij}. For power supply path $\forall\phi_{ij}\in\Phi_{ij}$, $\phi_{ij}=i\to\ldots\to m\to n\to\ldots\to j-1\to j$: if $E_m<E_n$, ϕ_{ij} is reasonable and $t_{ij}^T=1$; otherwise, ϕ_{ij} is unreasonable and $t_{ij}^T=0$.

The power supply rationality η_T^2 is expressed as:

$$\eta_T^2=\begin{cases}1 & \forall i,j\in\mathcal{V},t_{ij}^T=1\\ 0 & \exists i,j\in\mathcal{V},t_{ij}^T=0\end{cases} \tag{15}$$

In summary, only when $\eta_T^1=1$ and $\eta_T^2=1$ are satisfied at the same time can $\mathcal{G}_T$ satisfy the structural rationality in the Ω_T generation process.

D. Operational Rationality Calculation

This subsection presents the solution for rationality I_T^1 (corresponding to ③ in *Step 2* in Section II-B). For $\forall l\in\mathcal{C}_T$ in $\mathcal{G}_T$, an attacker can calculate the geographic length of cable l and estimate its impedance based on its potential model. The optimal power flow (OPF) [37], [38] is further calculated, and the objective is to minimize load shedding.

When no load shedding occurs in $\mathcal{G}_T$, the active line losses L_{ij} [39] are calculated as (16) to determine whether $\mathcal{G}_T$ can achieve the optimal operational efficiency and meet the operation economy as much as possible.

$$L_{ij}=\frac{r_{ij}x_{ij}^2}{r_{ij}+x_{ij}^2}p_{ij}^2H_B \tag{16}$$

where r_{ij} and x_{ij} are the resistance and reactance of cable l_{ij}, respectively; p_{ij} is the active power of cable l_{ij}; and H_B is the base value of line capacity.

The derivation of (16) can be found in Supplementary Material B. Therefore, the total active power loss of $\mathcal{G}_T$ is $\Delta P_T=\sum_{l_{ij}\in\mathcal{L}_T}L_{ij}$, where $\mathcal{L}_T=\mathcal{L}\bigcup\mathcal{C}_T$. I_T^1 is calculated by the total active power loss based on the DC power flow as:

$$I_T^1=\mathrm{e}^{-\Delta P_T} \tag{17}$$

The larger I_T^1 is, the more reasonable the proof $\mathcal{G}_T$ is, and the more likely it is close to $\mathcal{G}_0$.

IV. Stochastic Optimization Model for Vulnerability Identification Under Malicious Attacks

A. Probability Allocation Strategy

Based on Sections II-B and III, we obtain a topology set Ω_T with a quantity of N_W. For each topology $\mathcal{G}_T\in\Omega_T$, when the structural rationality, operational rationality, and complete cable rationality are satisfied (i. e., $\eta_T^1=\eta_T^2=1,I_T^1>\lambda_1$, and $I_T^2>\lambda_2$), we have $R_T>\xi$.

Taking Ω_T as the topology scenario for vulnerability identification, probabilities are allocated to each $\mathcal{G}_T$ based on R_T [30], [40]. Given that adding one more cable introduces additional errors, for $\mathcal{G}_T\in\Omega_T$, the scenario probability π_T of $\mathcal{G}_T$ is related to R_T, NL_T, and the probability of other topologies $\mathcal{G}_U$ in Ω_T as:

$$\pi_T=\frac{\frac{R_T}{NL_T}}{\sum_{\mathcal{G}_U\in\Omega_T}\frac{R_U}{NL_U}} \tag{18}$$

where R_U is the rationality of topology $\mathcal{G}_U$; and NL_U is the

number of cables in the completion cable set $\mathcal{C}_U$.

From (18), the scenario probability π_T of $\mathcal{G}_T$ increases with smaller NL_T and higher rationality index R_T.

B. Bi-level Stochastic Optimization Model

In this subsection, a bi-level stochastic optimization model is developed based on the extracted topology set Ω_T to identify the vulnerable nodes of UPN under malicious attacks, aiming to address the impact of topology completion uncertainty on the accuracy of vulnerability identification.

Based on the traditional stochastic programming formulation [41], [42], the objective function of the upper-level problem is to maximize the expected load shedding risk from optimal attack strategies in $N-K$ attack scenarios, as given in (19).

$$\max_{a} \sum_{\mathcal{G}_T \in \Omega_T} \pi_T \left(\sum_{i \in \mathcal{V}} \Delta Pd_{i*}^{T} \right) \tag{19}$$

where ΔPd_{i*}^{T} is the load shedding of node i caused by the attack strategy in $\mathcal{G}_T$, and it is determined by the optimal load shedding result obtained from the lower-level optimization problem.

$$\sum_{i \in \mathcal{V}} \Delta Pd_{i*}^{T} = \arg\min \sum_{i \in \mathcal{V}} \Delta Pd_{i}^{T} \tag{20}$$

For the development of attack strategy, let a_i be the binary variable to represent if node i is under attack: if yes, $a_i=1$; otherwise, $a_i=0$. Let K be the total number of attacked nodes and $\mathcal{N}_a=\{i|a_i=1,\forall i\in\mathcal{V}\}$ represent the set of all attacked nodes, i.e., the attack strategy.

Therefore, the main constraint of the upper-level problem is formulated as follows. The attacker selects power nodes for targeted attacks under limited attack resources, and the total number of attacked nodes in $\mathcal{N}_a$ is K. The constraint can be expressed as:

$$\sum_{i \in \mathcal{V}} a_i = K \tag{21}$$

Other constraints of the upper-level problem are used to calculate the state of lines in $\mathcal{G}_T$ after attack.

Equation (22) indicates the state of an attacked node i ($i \in \mathcal{N}_a$). Equation (23) indicates that the state of non-attacked node ($i \notin \mathcal{N}_a$) depends on that of its adjacent nodes: if all adjacent nodes fail, $\tau_i=0$; otherwise, $\tau_i=1$.

$$\tau_i = 1 - a_i \quad \forall i \in \mathcal{N}_a \tag{22}$$

$$\tau_i = \begin{cases} 0 & \sum_{j \in \mathcal{N}_i^T} \tau_j = 0, \forall i \notin \mathcal{N}_a \\ 1 & \sum_{j \in \mathcal{N}_i^T} \tau_j > 0, \forall i \notin \mathcal{N}_a \end{cases} \tag{23}$$

where $\mathcal{N}_i^T$ is the set of adjacent nodes for node i in $\mathcal{G}_T$.

Equation (24) is the constraint of cable state in $\mathcal{G}_T$ after attack. The state of cable l_{ij} v_{ij}^T is determined by the states of nodes i and j at the beginning and end. If $\tau_i=1$ and $\tau_j=1$, $v_{ij}^T=1$; otherwise, $v_{ij}^T=0$.

$$v_{ij}^T = \begin{cases} 1 & \tau_i+\tau_j=2, \forall l_{ij} \in \mathcal{L}_T \\ 0 & \tau_i+\tau_j<2, \forall l_{ij} \in \mathcal{L}_T \end{cases} \tag{24}$$

Note that (23) and (24) can be equivalently converted to constraints (25) and (26), respectively, by the big-M method [43].

$$0 \le \tau_i \le M \sum_{r \in \mathcal{N}_i^T} \tau_r \quad \forall i \notin \mathcal{N}_a \tag{25}$$

$$\begin{cases} 0 \le 1 - v_{ij}^T \le \left(2 - \tau_i - \tau_j\right) M \\ 0 \le v_{ij}^T \le \tau_i M \\ 0 \le v_{ij}^T \le \tau_j M \end{cases} \quad \forall l_{ij} \in \mathcal{L}_T \tag{26}$$

In the lower-level problem, the system operator solves a DC OPF in each topology $\mathcal{G}_T$ as:

$$\min \sum_{i \in \mathcal{V}} \Delta Pd_i^T \tag{27}$$

s.t.

$$p_{ij}^T = v_{ij}^T \frac{1}{x_{ij}} \left(\delta_i^T - \delta_j^T\right) \quad l_{ij} \in \mathcal{L}_T \tag{28}$$

$$\sum_{k \in \mathcal{V}_{G(i)}} g_k^T - \sum_{l_{ij} \in \mathcal{L}_T} A_{ij}^T p_{ij}^T + \Delta Pd_i^T = Pd_i \quad i \in \mathcal{V} \tag{29}$$

$$\underline{p}_{ij} \le p_{ij}^T \le \bar{p}_{ij} \quad l_{ij} \in \mathcal{L}_T \tag{30}$$

$$\underline{g}_k \le g_k^T \le \bar{g}_k \quad k \in \mathcal{V}_G \tag{31}$$

$$0 \le \Delta Pd_i^T \le Pd_i \quad i \in \mathcal{V} \tag{32}$$

The objective of the lower-level problem in (27) is to minimize the total load shedding under the failed lines in topology $\mathcal{G}_T$, determined by the attacker in the upper-level problem. Constraint (28) is the power flow equation, indicating the power flow p_{ij}^T of cable l_{ij} is affected by its state v_{ij}^T in $\mathcal{G}_T$ after attack, determined by the upper-level problem, where δ_i^T and δ_j^T are the nodal phase angles of nodes i and j, respectively. Constraint (29) is the nodal power balance equation, where g_k^T is the generator power output in $\mathcal{G}_T$ after attack; $\mathcal{V}_{G(i)}$ is the generator set of node i, Pd_i is the load of node i; and A_{ij}^T is the element of the network incidence matrix: $A_{ij}^T=1$ if node i is the sending node of line l_{ij}, $A_{ij}^T=-1$ if node i is the receiving node of line l_{ij}, and $A_{ij}^T=0$ otherwise. Constraint (30) is the corresponding line flow capacity limits, where $\overline{(\cdot)}$ and $\underline{(\cdot)}$ are the upper and lower limits, respectively. Constraints (31) and (32) are the generator output limit and load shedding limit, respectively.

C. Equivalent S-MILP Model

To address the nonlinearity caused by the multiplication of integer and continuous variables in (28), we first linearize it using the method from [17]. Subsequently, the lower-level model described in (27)-(32) is reformulated as constraints for the upper-level model using KKT conditions. As a result, the original model is converted into an S-MILP one, which is expressed as:

$$\max_{a} \sum_{\mathcal{G}_T \in \Omega_T} \pi_T \sum_{i \in \mathcal{V}} \Delta Pd_i^T \tag{33}$$

s.t.

$$\text{(21)-(24), (29)-(32)}$$

$$p_{ij}^T = \frac{1}{x_{ij}} \left(z_i^T - z_j^T\right) \quad l_{ij} \in \mathcal{L}_T \tag{34}$$

$$z_i^T = \delta_i^T - s_i^T \quad l_{ij} \in \mathcal{L}_T \tag{35}$$

$$z_j^T=\delta_j^T-s_j^T \quad l_{ij}\in\mathcal{L}_T \tag{36}$$

$$\underline{\delta}\,v_{ij}^T\le z_i^T\le\overline{\delta}\,v_{ij}^T \quad l_{ij}\in\mathcal{L}_T \tag{37}$$

$$\underline{\delta}\,v_{ij}^T\le z_j^T\le\overline{\delta}\,v_{ij}^T \quad l_{ij}\in\mathcal{L}_T \tag{38}$$

$$\underline{\delta}\left(1-v_{ij}^T\right)\le z_i^T\le\overline{\delta}\left(1-v_{ij}^T\right) \quad l_{ij}\in\mathcal{L}_T \tag{39}$$

$$\underline{\delta}\left(1-v_{ij}^T\right)\le z_j^T\le\overline{\delta}\left(1-v_{ij}^T\right) \quad l_{ij}\in\mathcal{L}_T \tag{40}$$

$$\sum_{l_{ij}\in\mathcal{L}_T}\frac{1}{x_{ij}}A_{ij}^Tt_{ij}^T=0 \quad i\in\mathcal{V} \tag{41}$$

$$t_{ij}^T=\mu_{ij}^T-h_{ij}^T \quad l_{ij}\in\mathcal{L}_T \tag{42}$$

$$\underline{\mu}_{ij}v_{ij}^T\le t_{ij}^T\le\overline{\mu}_{ij}v_{ij}^T \quad l_{ij}\in\mathcal{L}_T \tag{43}$$

$$\underline{\mu}_{ij}\left(1-v_{ij}^T\right)\le h_{ij}^T\le\overline{\mu}_{ij}\left(1-v_{ij}^T\right) \quad l_{ij}\in\mathcal{L}_T \tag{44}$$

$$-\lambda_i^T-\underline{\theta}_k^T+\overline{\theta}_k^T=0 \quad k\in\mathcal{V}_G \tag{45}$$

$$\lambda_i^T-\lambda_j^T-\mu_{ij}^T-\underline{\omega}_{ij}^T+\overline{\omega}_{ij}^T=0 \quad l_{ij}\in\mathcal{L}_T \tag{46}$$

$$1-\lambda_i^T-\underline{\alpha}_i^T+\overline{\alpha}_i^T=0 \quad i\in\mathcal{V} \tag{47}$$

$$\underline{\omega}_{ij}^T\ge 0 \quad l_{ij}\in\mathcal{L}_T \tag{48}$$

$$\overline{\omega}_{ij}^T\ge 0 \quad l_{ij}\in\mathcal{L}_T \tag{49}$$

$$\underline{\theta}_k^T\ge 0 \quad k\in\mathcal{V}_G \tag{50}$$

$$\overline{\theta}_k^T\ge 0 \quad k\in\mathcal{V}_G \tag{51}$$

$$\underline{\alpha}_i^T\ge 0 \quad i\in\mathcal{V} \tag{52}$$

$$\overline{\alpha}_i^T\ge 0 \quad i\in\mathcal{V} \tag{53}$$

$$\underline{\omega}_{ij}^T\le M\omega_{ij}^{\underline{\omega}T} \quad l_{ij}\in\mathcal{L}_T \tag{54}$$

$$p_{ij}^T+\overline{p}_{ij}\le M\left(1-\omega_{ij}^{\underline{\omega}T}\right) \quad l_{ij}\in\mathcal{L}_T \tag{55}$$

$$\overline{\omega}_{ij}^T\le M\omega_{ij}^{\overline{\omega}T} \quad l_{ij}\in\mathcal{L}_T \tag{56}$$

$$\underline{p}_{ij}-p_{ij}^T\le M\left(1-\omega_{ij}^{\overline{\omega}T}\right) \quad l_{ij}\in\mathcal{L}_T \tag{57}$$

$$\underline{\theta}_k^T\le M\omega_k^{\underline{\theta}T} \quad k\in\mathcal{V}_G \tag{58}$$

$$g_k^T-\underline{g}_k\le M\left(1-\omega_k^{\underline{\theta}T}\right) \quad k\in\mathcal{V}_G \tag{59}$$

$$\overline{\theta}_k^T\le M\omega_k^{\overline{\theta}T} \quad k\in\mathcal{V}_G \tag{60}$$

$$\overline{g}_k-g_k^T\le M\left(1-\omega_k^{\overline{\theta}T}\right) \quad k\in\mathcal{V}_G \tag{61}$$

$$\underline{\alpha}_i^T\le M\omega_i^{\underline{\alpha}T} \quad i\in\mathcal{V} \tag{62}$$

$$\Delta Pd_i^T\le M\left(1-\omega_i^{\underline{\alpha}T}\right) \quad i\in\mathcal{V} \tag{63}$$

$$\overline{\alpha}_i^T\le M\omega_i^{\overline{\alpha}T} \quad i\in\mathcal{V} \tag{64}$$

$$Pd_i-\Delta Pd_i^T\le M\left(1-\omega_i^{\overline{\alpha}T}\right) \quad i\in\mathcal{V} \tag{65}$$

$$\omega_{ij}^{\underline{\omega}T}+\omega_{ij}^{\overline{\omega}T}\le 1 \quad l_{ij}\in\mathcal{L}_T \tag{66}$$

$$\omega_k^{\underline{\theta}T}+\omega_k^{\overline{\theta}T}\le 1 \quad k\in\mathcal{V}_G \tag{67}$$

$$\omega_i^{\underline{\alpha}T}+\omega_i^{\overline{\alpha}T}\le 1 \quad i\in\mathcal{V} \tag{68}$$

Constraints (34) - (40) are the linearized relaxation constraints of (28) by introducing four continuous variables $z_i^T=v_{ij}^T\delta_i^T$, $z_j^T=v_{ij}^T\delta_j^T$, s_i^T, and s_j^T, where $\underline{\delta}$ and $\overline{\delta}$ are the lower bound and upper bound for the nodal phase angles, respectively. Constraints (41) - (53) are the linearized dual constraints of the original lower-level problem, where μ_{ij}^T, λ_i^T, $\overline{\omega}_{ij}^T$, $\underline{\omega}_{ij}^T$, $\overline{\theta}_k^T$, $\underline{\theta}_k^T$, $\overline{\alpha}_i^T$, and $\underline{\alpha}_i^T$ are the Lagrange multipliers related to DC power flow constraints (28)-(32); and t_{ij}^T and h_{ij}^T are the intermediate variables introduced to linearize the dual constraints. Constraints (54) - (68) represent the linearized complementary relaxation constraints, where $\omega_{ij}^{\underline{\omega}T}$, $\omega_{ij}^{\overline{\omega}T}$, $\omega_k^{\underline{\theta}T}$, $\omega_k^{\overline{\theta}T}$, $\omega_i^{\underline{\alpha}T}$, and $\omega_i^{\overline{\alpha}T}$ are the 0-1 variables used to linearize the original complementary slackness constraints.

The linearization process for (28) and the detailed KKT transformation of the lower-level problem are provided in Supplementary Material C.

V. Case Study

This section performs numerical simulations in an actual UPN to validate the effectiveness of UPN topology completion and vulnerability identification. Related calculations are performed in MATLAB using Yalmip programming. The solver utilizes CPLEX 12.8, with a computer setup featuring an Intel Core i7-10870H CPU (2.21 GHz) and 16 GB of memory.

A. Accuracy Verification of I-SHRINK Algorithm

As shown in Fig. 5, the UPN topology $\mathcal{G}_0$ in this case consists of 29 substations (S_1-S_{29}) and 10 power plants (G_{30}-G_{39}). The visible overhead lines are represented by solid lines, while the cables that require completion are indicated by dashed lines. After removing the hidden lines, the partition result of $\mathcal{G}_0$ utilizing the I-SHRINK algorithm is illustrated in Fig. 5, with the colors signifying their memberships in the same partition.

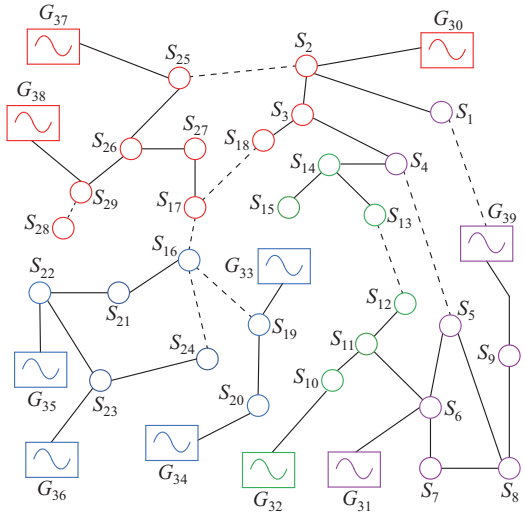


Fig. 5. Partitioning result of targeted incomplete UPN topology $\mathcal{G}_0$.

Supplementary Material D details the relationship between $\mathcal{G}_0$ and road networks, including information on cable laying index $\mathbb{E}_l$, various classes of cable signs (A, B, C), road

lengths, and power system parameters.

By taking $\xi=0.63$ as the value of pre-set threshold for rationality filter, 3000 topologies are selected by Monte Carlo sampling to construct the complete topology set Ω_T ($R_T>\xi$) following the steps in Section II-B. The three rationality metrics for $\mathcal{G}_0$, i.e., structural rationality, operational rationality, and complete cable rationality, are 1, 0.6499, and 0.9706, respectively, resulting in rationality $R_{T0}=0.6308$.

Figure 6 shows a complete topology $\mathcal{G}_{T1}$ from Ω_T with a similar high rationality $R_{T1}=0.63$, where blue, purple, and red dashed lines represent the successfully, unsuccessfully, and wrong complete cables, respectively.

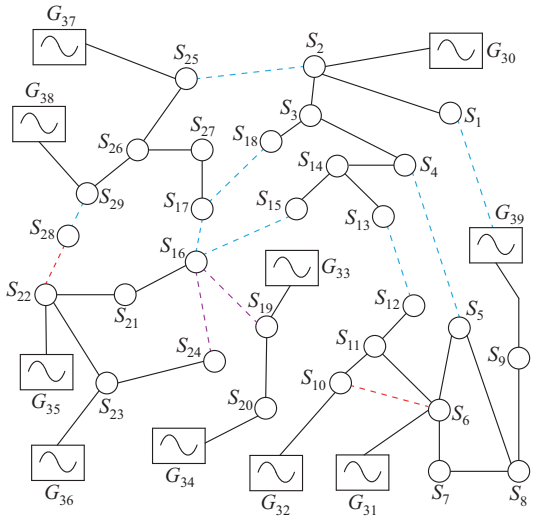


Fig. 6. A complete topology $\mathcal{G}_{T1}$ with high rationality.

The completion accuracy D_H of $\mathcal{G}_T$ is evaluated by the vector distance raised in [44] as:

$$D_H=\frac{1}{100}\left(100\frac{n_a}{n_0}+\frac{n_{af}}{n_0+\varepsilon_B}\right) \tag{69}$$

where n_a, n_0, and n_{af} are the number of successfully complete cables, the number of total invisible cables to be complete, and the number of unsuccessfully complete cables, respectively; and ε_B is a small positive constant. For $\mathcal{G}_{T1}$, $D_H\approx 0.67$.

We consider the complete topology to be relatively accurate when about half of the cables are successfully completed, i.e., $D_H\approx 0.5$. Therefore, with different ξ, the Monte Carlo sampling is used to obtain different Ω_T ($R_T>\xi$). By calculating the average value of D_H in Ω_T ($R_T>\xi$), i.e., $\bar{D}_H$, it can be observed that $\bar{D}_H\approx 0.5$ when $\xi=0.46$.

The curve of $\bar{D}_H$ with changes in ξ is shown in Fig. 7. It is important to highlight that $R_{T0}=0.6308$. However, as illustrated in Fig. 7, there exist complete topologies with $R_T>R_{T0}$. This situation is likely to occur as $\mathcal{G}_0$ may not always be the optimal completion solution. For instance, taking $\xi=0.64>R_{T0}$ as a case study, we extract 10 complete topologies with $R_T>R_{T0}$ and present their rationality metrics in Table II.

Table II shows that, while structural rationality is maintained, the operational rationalities of $\mathcal{G}_{T2}$-$\mathcal{G}_{T6}$ exceed that of $\mathcal{G}_0$. This indicates the potential for optimizing the operational efficiency and power supply capacity in $\mathcal{G}_0$. Additionally, the complete cable rationalities of $\mathcal{G}_{T7}$-$\mathcal{G}_{T9}$ exceed that of $\mathcal{G}_0$, suggesting that attackers with knowledge of road and cable signs can obtain more rational complete cables than $\mathcal{G}_0$. This highlights the opportunities for optimizing the current layout and trajectory of cables. When $\xi=0.64$, there are even completion topologies $\mathcal{G}_{T10}$ and $\mathcal{G}_{T11}$, whose operational rationalities and complete cable rationalities are higher than those of $\mathcal{G}_0$.

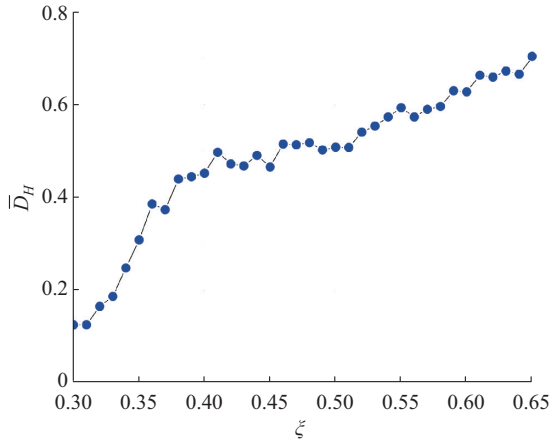


Fig. 7. Curve of $\bar{D}_H$ with changes in ξ.

TABLE II
RATIONALITY METRICS OF COMPLETE TOPOLOGY EXTRACTED WITH $R_T>R_{T0}$

Topology	$\eta_T^1\eta_T^2$	I_T^1	I_T^2	R_T
$\mathcal{G}_{T2}$	1	0.97229	0.66257	0.64421
$\mathcal{G}_{T3}$	1	0.97203	0.66804	0.64936
$\mathcal{G}_{T4}$	1	0.97087	0.66204	0.64276
$\mathcal{G}_{T5}$	1	0.97077	0.66606	0.64659
$\mathcal{G}_{T6}$	1	0.97203	0.66804	0.64936
$\mathcal{G}_{T7}$	1	0.96868	0.67146	0.65043
$\mathcal{G}_{T8}$	1	0.96961	0.66307	0.64292
$\mathcal{G}_{T9}$	1	0.97045	0.66823	0.64848
$\mathcal{G}_{T10}$	1	0.97077	0.66172	0.64237
$\mathcal{G}_{T11}$	1	0.97121	0.66687	0.64767

The analysis above shows that the actual topology may not be the most rational one with the highest R_T. To effectively minimize the impact of topology completion uncertainty on vulnerability identification results, it is essential to consider various possible topologies with high R_T. This validates the significance of considering the topology completion uncertainty in the vulnerability identification model.

B. Vulnerable Node Identification

Two sets of topologies Ω_{T1} and Ω_{T2} are selected with $\xi_1=0.4310$ and $\xi_2=0.6210$, respectively. Using the bi-level stochastic programming model, we identify the vulnerable nodes in the K-node attack scenario by ranking the load shedding risk. The top five combinations of vulnerable

nodes are selected for Ω_{T1} and Ω_{T2}, and the results are compared with $\mathcal{G}_0$ in Table III. Table III shows that in $\mathcal{G}_0$, the top five vulnerable nodes in 1-node attack scenario are nodes 16, 19, 29, 38, and 39. These nodes also appear in the 2-node and 3-node attack scenarios. Node 16 is the most vulnerable, being located in a critical power transmission corridor; its failure would result in load shedding risk of 1184.4, which is about 24.1% of the total system load. Node 19 is moderately important, and its failure would cause generator at node 10 to go offline, resulting in an isolation of nodes 5 and 20 and load shedding risk of 1133. The failure of node 29, ranking third, would take nodes 28 and 38 offline, resulting in load shedding risk of 1014.4. Meanwhile, nodes 38 and 39, ranking fourth and fifth, have critical power plants.

TABLE III
COMPARISON OF VULNERABLE NODES IN Ω_{T1}, Ω_{T2}, AND $\mathcal{G}_0$

K	Ω_{T1}		Ω_{T2}		$\mathcal{G}_0$	
	Vulnerable node No.	Load shedding risk	Vulnerable node No.	Load shedding risk	Vulnerable node No.	Load shedding risk
1	33	904.4	29	947.91	16	1184.4
	38	884.4	38	945.76	19	1184.4
	39	822.4	33	904.40	29	1014.4
	35	704.4	16	898.96	38	1014.4
	32	704.4	19	893.36	39	952.4
2	33, 38	1734.4	29, 33	1805.91	19, 29	2014.4
	33, 39	1672.4	33, 38	1797.19	16, 38	2014.4
	38, 39	1652.4	19, 29	1734.40	16, 29	2014.4
	32, 33	1554.4	16, 29	1734.40	19, 29	2014.4
	33, 35	1554.4	29, 39	1713.76	19, 39	1952.4
3	29, 33, 39	2502.4	29, 33, 39	2502.40	19, 38, 39	2782.4
	33, 38, 39	2502.4	33, 38, 39	2502.40	16, 38, 39	2782.4
	16, 22, 29	2455.9	16, 22, 29	2455.90	16, 29, 39	2782.4
	16, 29, 35	2455.9	16, 29, 35	2455.90	19, 29, 39	2782.4
	19, 22, 29	2447.2	19, 22, 29	2447.20	19, 22, 38	2664.4

However, for Ω_{T1} and Ω_{T2}, since $\xi_1 < \xi_2$, the average values $\bar{R}_{T1} < \bar{R}_{T2}$ and $\bar{D}_{H1} < \bar{D}_{H2}$, indicating that the disparity between Ω_{T1} and $\mathcal{G}_0$ is greater than that between Ω_{T2} and $\mathcal{G}_0$. This discrepancy leads to a smaller proportion of successfully complete cables (e.g., 16-19, 16-15, 16-17) in $\mathcal{G}_0$ across all topologies in Ω_{T1} than that in Ω_{T2}. For example, the proportions of successfully complete cable 16-19 in Ω_{T1} and Ω_{T2} are 0.67 and 0.77, respectively. In the 1-node attack scenario, the lower proportion of successfully complete cable 16-19 in Ω_{T1} leads to nodes 16 and 19 receiving lower vulnerability rankings and not appearing in the top five positions. Conversely, in Ω_{T2}, nodes 16 and 19 rank the fourth and fifth, respectively, which bring the ranking of Ω_{T2} closer to that of $\mathcal{G}_0$. Similarly, in the 2-node and 3-node attack scenarios, topologies in Ω_{T2} include a greater number of complete cables connected to nodes 16 or 19, making their rankings of vulnerable node combinations more similar to those of $\mathcal{G}_0$.

At the same time, based on Table III, it is evident that in the 1-node, 2-node, and 3-node attack scenarios for Ω_{T1}, the proportion of generator nodes among the top-ranked vulnerable nodes is higher than that in $\mathcal{G}_0$, particularly nodes 33, 38, and 39. Although the substation at attack node 16 or 19 in $\mathcal{G}_0$ may cause more severe load shedding, attackers in Ω_{T1} and Ω_{T2} prefer targeting power plants due to their reliable impact. The low threshold ξ in Ω_{T1} restricts the precise completion of the topology, resulting in uncertainty regarding the effectiveness of attacks on substations. Therefore, a more reliable strategy is considered to target power plants to reduce power generation and cause load shedding.

For Ω_{T2}, the increase in the average rationality increases the confidence of attackers in the potential effects of targeting vulnerable nodes. Consequently, the vulnerability of nodes 16 and 19 increases significantly. In the 1-node, 2-node, and 3-node attack scenarios, the significance of node combinations that include nodes 16 and 19 increases, while power plants continue to dominate by occupying the top two positions. This highlights the necessity for attackers to utilize a topology that is reasonably rational or closely resembles G_0 to gather sufficient information for strategically targeting critical substation nodes and anticipating their returns.

C. Sensitivity Analysis

To further validate the conclusion from Section III-B, we describe the similarity degree τ_K between the ranking of vulnerable node combinations in Ω_T and $\mathcal{G}_0$:

$$\tau_K = 1 - \frac{1}{2}\frac{\sum_{i=1}^{R_K}\left|\omega_i^{o\text{-}K} - \omega_i^{\Omega\text{-}K}\right|}{R_K\left(R_K - 1\right)} \tag{70}$$

where R_K is the total number of possible vulnerable node combinations in the K-node attack scenario; and $\omega_i^{o\text{-}K}$ and $\omega_i^{\Omega\text{-}K}$ are the rankings of vulnerable node combinations in Ω_T and $\mathcal{G}_0$, respectively.

According to (70), the similarity degree τ_K ranges from 0

to 1, with τ_K closer to 1 indicating greater similarity. Figure 8 illustrates the curve of τ_K as ξ changes.

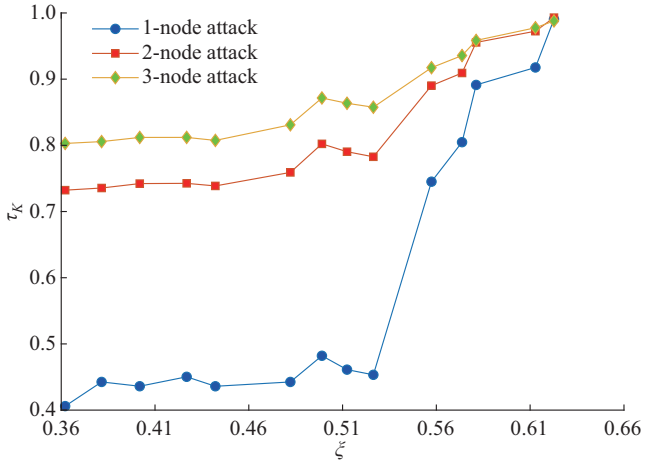


Fig. 8. Curve of τ_K as ξ changes.

As shown in Fig. 8, as ξ increases, the similarity degree τ_K between Ω_T and $\mathcal{G}_0$ gradually increases. Since $R_{T0}=0.64$, when $\xi>0.6$, the similarity degree τ_K exceeds 0.9. This means that the ranking using Ω_T with $\xi>0.6$ is very close to the real results, as if the attacker has access to the complete topology information of $\mathcal{G}$.

Based on the results above, we conclude that applying the three rationality metrics. i.e., structural rationality, operational rationality, and cable complete rationality, in Monte Carlo sampling will produce a highly rational and trustworthy Ω_T. More rational Ω_T enables the formulation of attack strategies that closely align vulnerability identification results with $\mathcal{G}_0$.

VI. Conclusion

This paper introduces an I-SHRINK algorithm considering structural rationality, operational rationality, and cable complete rationality for UPN topology completion. To reduce the impact of topology completion uncertainty on vulnerability identification of UPN, a set of topologies with high rationality is developed. Using this set of topologies, a bi-level stochastic optimization model is proposed to identify vulnerable nodes in $N-K$ attack scenarios.

This study shows that as the completion standard ξ increases, the average accuracy of topologies in Ω_T improves, indicating a greater similarity to $\mathcal{G}_0$ and validating the three rationality metrics. When comparing the rankings of vulnerable node combinations in K-node attack scenarios between Ω_T and $\mathcal{G}_0$ with different ξ, it is observed that attackers target key power plant nodes to maximize benefits at lower topology extraction criteria ξ. As ξ increases, the rankings of vulnerable node combinations in Ω_T become increasingly aligned with $\mathcal{G}_0$. Sensitivity analysis further validates this trend.

Based on the findings of vulnerability identification for UPN in this study, grid operators can strategically utilize information asymmetry to: ① conceal critical transmission lines, ② deploy deceptive line indicators, and ③ reconfigure power distribution networks to ensure the protection of vital loads. This comprehensive defense mechanism effectively misleads potential attackers by presenting them with manipulated topological information, thereby significantly increasing the complexity of constructing a coherent attack strategy. Consequently, this redirects adversarial attention toward less critical nodes, enhancing overall grid resilience. The development of robust defense strategies against incomplete information attacks will constitute a primary focus of our subsequent research endeavors. Another interesting area of research is the investigation of other incomplete UPN information, including generator parameters and loads.

References

[1] D. Du, M. Zhu, X. Li *et al.*, "A review on cybersecurity analysis, attackdetection, and attack defense methods in cyber-physical power systems," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 3, pp. 727-743, May 2023.

[2] G. Liang, S. R. Weller, J. Zhao *et al.*, "A framework for cyber-topology attacks: line-switching and new attack scenarios," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1704-1712, Mar. 2019.

[3] X. Huang, Z. Qin, M. Xie *et al.*, "Defense of massive false data injection attack via sparse attack points considering uncertain topological changes," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 6, pp. 1588-1598, Nov. 2022.

[4] R. Smith. (2009, Feb.). Assault on California power station raises alarm on potential for terrorism. [Online]. Available: https://www.wsj.com/articles/assault-on-california-power-station-raises-alarm-on-potential-for-terrorism-1391570879

[5] S. Yang, K. Lao, H. Hui *et al.*, "Secure distributed control for demand response in power systems against deception cyber-attacks with arbitrary patterns," *IEEE Transactions on Power Systems*, vol. 39, no. 6, pp. 7277-7290, Nov. 2024.

[6] Y. Xiang and L. Wang, "An improved defender-attacker-defender model for transmission line defense considering offensive resource uncertainties," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2534-2546, May 2019.

[7] G. Chen, Z. Y. Dong, D. J. Hill *et al.*, "Exploring reliable strategies for defending power systems against targeted attacks," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1000-1009, Aug. 2011.

[8] X. Dong, T. R. Nyberg, P. Hämäläinen *et al.*, "Vulnerability analysis of smart grid based on complex network theory," in *Proceedings of 2015 IEEE 5th International Conference on Information Science and Technology*, Changsha, China, Apr. 2015, pp. 525-529.

[9] E. I. Bilis, W. Kröger, and C. Nan, "Performance of electric power systems under physical malicious attacks," *IEEE Systems Journal*, vol. 7, no. 4, pp. 854-865, Dec. 2013.

[10] A. K. Srivastava, T. A. Ernster, R. Liu *et al.*, "Graph-theoretic algorithms for cyber-physical vulnerability analysis of power grid with incomplete information," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 5, pp. 887-899, Sept. 2018.

[11] Y. Zhu, J. Yan, Y. Tang *et al.*, "Resilience analysis of power grids under the sequential attack," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2340-2354, Dec. 2014.

[12] X. Wei, S. Gao, T. Huang *et al.*, "Complex network-based cascading faults graph for the analysis of transmission network vulnerability," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1265-1276, Mar. 2019.

[13] A. L. Motto, J. M. Arroyo, and F. D. Galiana, "A mixed-integer LP procedure for the analysis of electric grid security under disruptive threat," *IEEE Transactions on Power Systems*, vol. 20, no. 3, pp. 1357-1365, Aug. 2005.

[14] J. Salmeron, K. Wood, and R. Baldick, "Analysis of electric grid security under terrorist threat," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 905-912, May 2004.

[15] J. Salmeron and R. K. Wood, "The value of recovery transformers in protecting an electric transmission grid against attack," *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2396-2403, Sept. 2015.

[16] Z. Li, M. Shahidehpour, A. Alabdulwahab *et al.*, "Analyzing locally coordinated cyber-physical attacks for undetectable line outages," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 35-47, Jan. 2018.

[17] J. M. Arroyo and F. D. Galiana, "On the solution of the bilevel programming formulation of the terrorist threat problem," *IEEE Transac-*

tions on Power Systems, vol. 20, no. 2, pp. 789-797, May 2005.
[18] H. T. Reda, A. Anwar, A. Mahmood *et al*., "Data-driven approach for state prediction and detection of false data injection attacks in smart grid," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 2, pp. 455-467, Mar. 2023.
[19] Y. Song, X. Liu, Z. Li *et al*., "Intelligent data attacks against power systems using incomplete network information: a review," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 4, pp. 630-641, Jul. 2018.
[20] W. Ma, J. Fang, and J. Wu, "Analyzing robustness of complex networks against incomplete information," *IEEE Transactions on Circuits and Systems II*: *Express Briefs*, vol. 69, no. 5, pp. 2523-2527, May 2022.
[21] A. Srivastava, T. Morris, T. Ernster *et al*., "Modeling cyber-physical vulnerability of the smart grid with incomplete information," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 235-244, Mar. 2013.
[22] House of Lords Library. (2024, Feb.). Undergrounding electrical transmission cables. [Online]. Available: https://lordslibrary. parliament. uk/undergrounding-electrical-transmission-cables
[23] X. Yang, M. Sheng, Y. Yuan *et al*., "Network topology inference from heterogeneous incomplete graph signals," *IEEE Transactions on Signal Processing*, vol. 69, pp. 314-327, Dec. 2020.
[24] A. Buciulea, S. Rey, and A. G. Marques, "Learning graphs from smooth and graph-stationary signals with hidden variables," *IEEE Transactions on Signal and Information Processing Over Networks*, vol. 8, pp. 273-287, Mar. 2022.
[25] J. Huang, H. Aun, J. Han *et al*., "SHRINK: a structural clustering algorithm for detecting hierarchical communities in networks," in P*roceedings of the 19th ACM International Conference on Information and Knowledge Management*, Toronto, Canada, Oct. 2010, pp. 219-228.
[26] Z. Zhu, T. Hu, D. Wu *et al.*, "Topology sensing of FANET under missing data," *Computer Networks*, vol. 255, p. 110856, Dec. 2024.
[27] W. Wang, M. Wang, X. Han *et al*., "Distributionally robust transmission expansion planning considering uncertainty of contingency probability," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 4, pp. 894-901, Jul. 2022.
[28] Z. Shi, Y. Xu, D. Xie *et al*., "Optimal coordination of transportable power sources and repair crews for service restoration of distribution networks considering uncertainty of traffic congestion," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 1, pp. 189-201, Jan. 2024.
[29] Q. Sun, Z. Wu, W. Gu *et al*., "Multi-stage co-planning model for power distribution system and hydrogen energy system under uncertainties," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 1, pp. 80-93, Jan. 2023.
[30] M. Carrion, J. M. Arroyo, and N. Alguacil, "Vulnerability-constrained transmission expansion planning: a stochastic programming approach," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1436-1445, Nov. 2007.
[31] R. Guimerà and L. A. N. Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 433, pp. 895-900, Feb. 2005.
[32] M. Chen, K. Kuzmin, and B. K. Szymanski, "Community detection *via* maximization of modularity and its variants," *IEEE Transactions on Computational Social Systems*, vol. 1, no. 1, pp. 46-65, Mar. 2014.
[33] W. Lin, X. Kong, P. S. Yu *et al*, "Community detection in incomplete information networks," in *Proceedings of the 21st international conference on World WideWeb*, Lyon, France, Apr. 2012, pp. 341-350.
[34] R. Guimera and M. Salespardo, "Missing and spurious interactions and the reconstruction of complex networks," *Proceedings of the National Academy of Sciences*, vol. 106, no. 52, pp. 22073-22078, Dec. 2009.
[35] IEEE. (2021, Nov.). IEEE approved draft standard for DC power transmission and communication to DC loads. [Online]. Available: https://ieeexplore.ieee.org/document/9540638
[36] Y. Jia, C. Andrieu, R. J. Piechocki *et al*., "Depth-first and breadth-first search based multilevel SGA algorithms for near optimal symbol detection in MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 7, no. 3, pp. 1052-1061, Mar. 2008.
[37] L. Gan and S. H. Low, "Optimal power flow in direct current networks," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2892-2904, Nov. 2014.
[38] M. Niu, C. Wan, and Z. Xu, "A review on applications of heuristic optimization algorithms for optimal power flow in modern power systems," *Journal of Modern Power Systems and Clean Energy*, vol. 2, no. 4, pp. 289-297, Nov. 2014.
[39] B. Stott, J. Jardim, and O. Alsac, "DC power flow revisited," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1290-1300, Aug. 2009.
[40] J. M. Arroyo, N. Alguacil, and M. Carrión, "A risk-based approach for transmission network expansion planning under deliberate outages," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1759-1766, Aug. 2010.
[41] R. Jiang, J. Wang, M. Zhang *et al*., "Two-stage minimax regret robust unit commitment," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2271-2282, Aug. 2013.
[42] Q. Zheng, J. Wang, and A. Liu, "Stochastic optimization for unit commitment: a review," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1913-1924, Jul. 2015.
[43] S. Pineda and J. M. Morales, "Solving linear bilevel problems using big-*M*: not all that glitters is gold," *IEEE Transactions on Power System*, vol. 34, no. 3, pp. 2469-2471, May 2019.
[44] P. Jiang, S. Huang, and T. Zhang, "Asymmetric information in military microgrid confrontations: evaluation metric and influence analysis," *Energies*, vol. 13, no. 8, p. 1954, Apr. 2020.

Chengze Li received the B.S. degree in electrical engineering and automation from China University of Petroleum (East China), Qingdao, China, in 2020. He is pursuing the Ph.D. degree in the Department of Electrical and Electronic Engineering, North China Electric Power University, Beijing, China. His research interests include vulnerability assessment, security, and resilience in power system.

Wenxia Liu received the B.S. degree in radio technology from Nanjing University of Science and Technology, Nanjing, China, in 1990, and the M.S. and Ph.D. degrees in electrical engineering and automation from Northeast Electric Power University, Jilin, China, and North China Electric Power University, Beijing, China, in 1995 and 2009, respectively. She is currently a Professor in the School of Electrical and Electronic Engineering, North China Electric Power University. Her research interests include risk assessment in power system and security in cyber-physical power system.

Rui Cheng received the Ph.D. degree in electrical engineering from Iowa State University, Iowa, USA, in 2023. He is currently an Assistant Professor at North China Electric Power University, Beijing, China. His research interests include optimization and machine learning in distribution system, and power system vulnerability and reliability.

Yuze Yang received the B.S. degree in electrical engineering and automation from Qingdao University of Science and Technology, Qingdao, China, in 2020. He is pursuing the Ph.D. degree in the Department of Electrical and Electronic Engineering, North China Electric Power University, Beijing, China. His research interests include cyber-physical power system and power system vulnerability.