An Attention Mechanism Based Algorithm of Identifying Vulnerable Spots of the Internet Data Center Power Systems

Xinpei Liu, Chunjian Kang*, Zhuo Jiang and Tiezheng Zhao

National Computer Network Emergency Response Technical Team /Coordination Center of China, Beijing, China

*vithon@163.com

Abstract. With the rapid development of Mobile Internet and Big Data in recent years, the demand for data centers is increasing. The number of super large data centers (more than 1000 counters) has increased. This paper focuses on the distribution system, who is responsible for providing a continuous, stable and uninterrupted power supply for the data center. In the event of a large-scale power outage, all business in the data center will be interrupted. Therefore, identifying which electrical equipment in the data center play the key role in a large-scale power outage is extremely important. Considering the computational complexity of the algorithm, the validity of the vulnerable nodes, the intelligence of the algorithm, this paper proposes a two-layer fragile node identification method based on attention mechanism. The algorithm selects the topology structure and electrical parameters of the data center power distribution system, and considers the flow load of the nodes around each electrical device to form an expression vector. After the vulnerability ranking of the candidate node expression vectors, the k most vulnerable nodes are selected. The outline of the two-layer node identification method is as follows, the first layer calculates an alternative node sequence through a variety of graph-based fragile node identification methods, and the second layer recalibrates the expression vector of each candidate node based on the attention mechanism and then sorts all candidate nodes. Through the simulation test of a real data center power distribution system, we can see that the proposed algorithm in this paper identifies the effectiveness of the vulnerable nodes in the data center power distribution system.

1. Introduction

A data center is a place to host a variety of information systems, including IT equipment and power infrastructure. The power infrastructure plays an important role in providing power support for information systems, and its reliable operation is very important for the stable operation of the upper information system. Without the 7 x 24 hours guarantee of the Data center infrastructure, the information system will not be able to perform its functions, such as cloud computing, artificial intelligence, big data, the Internet of Things and so on. These infrastructures include power systems, cooling systems, lighting systems, data center infrastructure management systems, video surveillance systems, fire protection systems, and other auxiliary but very important systems. In these infrastructures, the power system is the most important. A typical data center power system is shown in Figure 1.
Data center power distribution systems seem to be stable and reliable, but sometimes seemingly insignificant accidents can trigger the collapse and paralysis of the entire distribution network. For example, on July 3, 2009, a power distribution failed in Fisher Square, Seattle, causing the paralysis of Payment Portal, Authorize.net, Geocaching.com Service, Dotster Domain Name Registration Service, Microsoft Bing Travel Service and dozens of other websites [1]. On 23 December 2015, someone made a surprise cyber-attack on three Ukrainian regional electric power distribution companies, which cause several hours’ power supply interruption [2]. Therefore, power systems are very important in the data center infrastructure. If certain critical electrical equipment is shut down, some critical information services will fail. If critical information services are used to enhance network security, the attack will not be only an economic issue. Therefore, how to avoid large-scale power outage [3] becomes a key issue in the IDC power system maintenance.

At present, most data center owners pay insufficient attention to the power distribution system, and believe that the power distribution system is stable and reliable. However, as the operating period of power equipment increases, the equipment ages seriously, the distribution lines are complicated. Failures in specific links may cause large-scale power outages.

At present, the following problems exist in the maintenance of the data center power distribution system:

(1) It is judged by experience which link is of high importance, and how much impact can be caused by a failure;

(2) When an important power distribution switch fails, the maintenance personnel need to find the historical load record and calculate it manually. Based on the calculation result, it is decided whether to contact the bus switch for power supply. This method delays the emergency response time, and may also expand the fault range due to calculation errors.

(3) Most data centers use indiscriminate maintenance mode or several important node maintenance modes when performing power distribution equipment maintenance. As the weak links in the distribution path cannot be accurately identified, the maintenance personnel can’t be able to use limited resources in the most vulnerable areas.

Therefore, finding the most vulnerable nodes of the distribution system, and then strengthening the routine maintenance of such nodes and pre-planning the emergency plan will greatly reduce the impact of the distribution system failure.

There are many definitions and frameworks of power system vulnerability [4]. To effectively quantify the vulnerability of data center power systems, we define the vulnerability of data center power systems as a measure of system vulnerability relative to a series of cascaded events. The vulnerable spots of the IDC power system are the devices which fail may result in large-scale power-
off loads. The power-down load can be evaluated by the following equation. The vulnerability of the node is evaluated by the amount of power loss after the node fails.

\[
\text{Loss} = \left(1 - \frac{\sum_{i \in T_{\pi} \cap \text{load}} \text{load}_i}{\sum_{j \in G \cap \text{load}} \text{load}_j}\right) \times 100\% \tag{1}
\]

Where Loss is the ratio of lost loads in the IDC power network when some devices fail, \(T_{\pi}\) is the stable operating topology of the IDC power network after some failure occurs. \(t_l\) is the subset of load nodes of \(T_{\pi}\), which involves all the working loads. \(G\) is the topology graph of the normal IDC power network without any failure. \(g_l\) is the subset of load nodes of \(G\), which involves all the working and nonworking loads. \(\text{load}_i\) is the power load of load \(i\).

We believe that the larger the power-off load caused by the failure of electrical devices, the weaker the electrical devices are in the IDC power system. If the exhaustive method is applied, assuming that an IDC power system has \(n\) devices, the time complexity of finding the maximum lost loads caused by \(k\) broken-down devices is \(O(C^n_k)\). The problem falls into the \(N-k\) problem \([4]\). For large data centers, this computational complexity is unacceptable. It is a NP-hard problem.

The main contributions of this paper are as follows:

1. This paper proposes a new two-layer algorithm BA (bilevel algorithm): the first layer calculates an alternative node sequence through a variety of graph-based fragile node identification methods, and the second layer calibrates the expression vector for each candidate node based on the attention mechanism. And then sort the nodes to get the node vulnerability sequence.

2. In the second layer of the BA algorithm, the expression vector is calibrated for each candidate node based on the attention mechanism, considering not only the topology of the data center power distribution system, electrical parameters, but also the flow load of the nodes around each electrical device.

The rest of this paper is organized as follows: Section 2 introduce the related works which involves the classification of blackouts and the methods to analyze the vulnerability of the power systems. In Section 3, build the model of the data center power network. In Section 4, we propose a bilevel algorithm based on attention mechanism to identify the vulnerable spots of the data center power network. In Section 5, we present the experimental results. Finally, conclusions are provided in Section 6.

2. Related works
Since almost all the devices in the data center are powered by electricity, the impacts are tremendous when the power system outage \([5]\). There are mainly three categories of power system failures, which are natural hazards \([6,7]\), random failures\([8,9]\), intentional attacks \([10,11,12,13,14,15]\).

Many strategies can be applied to achieve the malicious purpose. Paul Cuffe, etc. compared eight attack strategies in the paper \([10]\), which are standard genetic algorithm, method of minimized fitness function by link survivability\([15]\), mixed integer linear programming method, random method, removing the most-heavily loaded \(K\) lines, electrical betweenness method, betweenness centrality method and topological metric of edge range method. Details can refer to the paper \([10]\).

On the vulnerability analysis of the power system, a lot of excellent results have been achieved in the above literature. Through analysis, there are mainly static, dynamic, and synthetic methods for fault outage modeling of the power system. The methods of vulnerability analysis mainly fall into heuristic Complex network methods, exhaustive search, engineering expertise, and artificial intelligence.

The above methods have achieved remarkable results, but they are either computationally complex or carefully tuned. Different from the above method, this paper proposes a new two-layer algorithm BA.
3. Model establishment

3.1. Power distribution system modeling

Large-scale data center power distribution system consists of power grid, high-voltage incoming line cabinet, isolation cabinet, metering cabinet, high-voltage feeder cabinet, transformer, busbar, low-voltage inlet cabinet, low-voltage output cabinet, reactive power compensation cabinet, connecting cable, secondary power distribution cabinets, UPS, three-level power distribution cabinets, various loads (servers, water-cooled units, pumps, lighting, office loads, air conditioning loads, compressors, and other motor-type loads). The above electrical equipment can be mainly divided into substation equipment (transformers, etc.), power transmission equipment (such as various types of power distribution cabinets), electrical equipment (such as various types of loads), and various types of electrical equipment are connected by cable or bus connection to transmit power. Moreover, the power transmission direction is transmitted only from the power transmitting end to the power receiving end.

The topology of a large-scale data center power distribution system can be represented by the weighted directed graph $G = (V, E)$, $V$ is a vertex set and the abstract modeling of various types of electrical equipment, including power distribution cabinets, transformers, tie switches, ats, sts, mts, ups, diesel generators, disconnectors, etc.; $E \subseteq V \times V$ is a set of directed edges. Elements $e = (x, y) \in E$ means the connecting electrical line of the two devices $x$ and $y$, including the transmission cable, the transmission bus, etc., the capacity (maximum supported electric power transmission capability) is denoted by $C(e)$. There are three types of vertices in the vertex set $V$ of the data center distribution network: substation equipment, power transmission equipment, and power equipment. Modeling the power distribution system to the network [16] software as $\tilde{G} = (V, E)$The generated model diagram is shown in Figure 2. After statistics, one data center has deployed 6 2500kVA transformers with a rated capacity of 15MW, equivalent to 4,500 ordinary household electric power, with an average annual power consumption of 20 million degrees, and 5 sets of 1250kw cooling water-cooled units, 12 sets of 400kVAUPS, and 1920 units. Lead-acid batteries, 580 cabinets, tens of orders of moving ring monitoring points and countless office outlets, lighting, etc., these power 24h uninterrupted support for the stable operation of information services. The modeled graph is shown in Figure 2.

![Figure 2](image)

Figure 2. Abstract modeling diagram of the distribution system.

3.2. A Fragile Node Identification Method Based on Attention Mechanism

Wherever possible try to ensure that the size of the text in your figures (apart from superscripts/subscripts) is approximately the same size as the main text (11 points). It is not comprehensive enough that the above various central calculation methods are considered from a
certain aspect. Therefore, a graph based two-layer fragile node identification method $BA$ is proposed. First, the above four central methods are adopted in the first layer. Calculate a vulnerable node candidate sequence $VB$, the formula is as follows.

- $VB = \{B_D, B_C, B_B, B_E\}$
- $B_D = [v_D^1, v_D^2, ..., v_D^k]$
- $B_C = [v_C^1, v_C^2, ..., v_C^k]$
- $B_B = [v_B^1, v_B^2, ..., v_B^k]$
- $B_E = [v_E^1, v_E^2, ..., v_E^k]$

In each of the central candidate sequences $B$, there are $k$ ($k = 10$ in this case) node sequences, and the nodes in different candidate sequences $B$ may be identical. Secondly, in the second layer, this paper proposes a novel node vulnerability identification method based on the attention mechanism. The idea of this algorithm is: the vulnerability of each node is not only related to its own attributes, such as the rated value of the circuit breaker, the importance of the node in the topology of the power distribution system, etc., but also related to the load condition of this node. Divided into 5 steps.

(1) The first step is to calculate the proportion of each node in each central sequence, and the calculation formula is as follows

$$bv_F^i = \frac{C_F(i)}{\sum_{j=1}^{k} C_F(j)}$$

(2) The second step is to calculate the load condition of each node. The statistical method is to calculate the sum of the load of the node. If one node has one parent node (for example, $f_s$), the load is recorded as $1/f_s$, as follows

$$l_i = \sum_{j=1}^{m} \frac{l_j}{f_{s_j}}$$

(3) The third step is to construct the expression vector of each node $rv_i$, this vector is rated by the node $e$, each central value $bv_F^i$ and the load situation $l_i$ and so on.

$$rv_i = [e, bv_D^i, bv_B^i, bv_C^i, bv_E^i, l_i]$$

(4) The fourth step is to map the expression vector of node $i$ into a quantizable node characteristic $vv_i$, calculated as follows

$$vv_i = 1/(1 + \exp (- \sum_{j=1}^{6} w_j \cdot rv_j^i)) \cdot rv_j^i \in rv_i$$

(5) Finally, the above four steps of each node in the candidate sequence $VB$ are sequentially calculated from the end load layer to the power output end layer, and the characteristics of each node are obtained $vv_i$. And the nodes in the $VB$ sequence are sorted according to the node characteristics from large to small, and the largest $k$ nodes are selected to form a sequence set $vul = \{v^1_y, v^2_y, ..., v^k_y\}$.

The algorithm block diagram is as the following Figure 3.
The pseudo-code of BA algorithm is as follows:

| Algorithm: BA (An Attention-base Bilevel algorithm of identifying the vulnerable spots of IDC power system) |
| --- |
| // Return a sequential vulnerable nodes’ set VUL |
| Require: G (The graph of the power network), Model DCNPDEM |
| // Return VUL (The vulnerable nodes) |
| VB < −∅ |
| for m in {B_D, B_C, B_B, B_E} do: |
| VB.append (Centrality (m)) |
| for i in VB do: |
| rv_i = [e, bv_D, bv_B, bv_C, bv_E, l_i] |
| vv_i = f(rv_i) |
| VUL = Sort (vv_i) |

4. Simulation and experiment

The experiment in this paper is based on a real data center in Beijing. After modeling, the data center distribution system has 594 nodes, 719 connection edges, 54 information system load columns, and 151 other types of air conditioners. This paper uses Python3 and graph theory and complex network modeling software package networkx to model and simulate the distribution network of a large data center. The graph model of this power distribution system is shown in Figure 2.

Some nodes are tripped by using degree centrality, Closeness centrality, betweenness centrality, eigenvector centrality, BA algorithm and random method. The trip strategy is as follows:

1. Random trip node: randomly trip 1 node at a time, and gradually increase the number of tripped nodes, and take the average value of 10 simulation results to form a curve;
2. The first k central nodes with the four centralities are tripped in turn, and the number of nodes increased;
3. Trip from the beginning to the end in the sequence of vulnerable nodes derived from the BA algorithm of this paper,
4. Under the six-node trip strategy, Figure 4 shows the relationship between the number of initial fault nodes of the power distribution system and the load shedding situation.
Selective tripping of some nodes has a greater impact on the power distribution system. Two (0.34%) tripped nodes may cause more than 20% loss of the load, indicating that there is a weak link in the power distribution system;

(2) When 10 nodes are tripped in the sequence of vulnerable nodes derived from the BA algorithm, it will cause 100% load loss of the power distribution network, indicating that the BA algorithm can effectively capture the vulnerable nodes of the data center power distribution system;

(3) When there are less than 3 faulty nodes, that is, when the power distribution system does not have a large-scale power outage fault, the BA algorithm and degree centrality can accurately grasp which faults have a greater impact on the power distribution system. This is because some electrical equipment in the distribution system has a large number of loads and load heavy, such as some ATS, UPS, etc.

(4) When there are more than 3 fault nodes and less than 7, the BA algorithm and the betweenness centrality can accurately grasp which faults have a greater impact on the power distribution system. This is because some electrical equipment of the data center power distribution system acts as a bridge for energy transmission. It is also a bottleneck in the data center power distribution system, such as some main switch, tie switch, etc.

5. Conclusion
Through the analysis of this paper, it can be seen that the distribution system does have vulnerable nodes. After such a node failure, it will lead to large-scale power outage, and thus the information service is affected. This paper proposes a two-layer fragile node identification method based on attention mechanism. It not only considers the topology structure and electrical parameters of the data center power distribution system, but also considers the flow load of nodes around each electrical device to form an expression vector. After sorting the candidate nodes, the k most vulnerable nodes are selected. Through the simulation test of a real data center distribution system, we can see that the BA algorithm in this paper can effective identify the of the vulnerable nodes in the data center power distribution system.

References
[1] Zhongguancun online enterprise station China. Fire attacks of the data center; 2012. [Online]. Available: http://server.zol.com.cn/337/3371759_all.html; [accessed 20 August 2019].

[2] Gaoqi Liang, Steven R. Weller, Junhua Zhao, Fengji Luo, Zhao Yang Dong The 2015 ukraine blackout: Implications for false data injection attacks IEEE Transactions on Power Systems, 32 (2017), pp. 3317-3318

[3] Lucas Cuadra, Sancho Salcedo-Sanz, Javier Del Ser, Silvia Jiménez-Fernández, Zong Woo Geem A Critical Review of Robustness in Power Grids Using Complex Networks Concepts Energies, 8
Caplar R and Kulisic P 1973 Proc. Int. Conf. on Nuclear Physics (Munich) vol 1 (Amsterdam: North-Holland/American Elsevier) p 517

[4] Amin Abedi, Ludovic Gaudard, Franco Romerio Review of major approaches to analyze vulnerability in power system Reliability Engineering & System Safety, 183 (2019), pp. 153–172

[5] Francisco Gutierrez, E. Barocio, F. Uribe, P. Zuniga Vulnerability Analysis of Power Grids Using Modified Centrality Measures Discrete Dynamics in Nature and Society, (2013)

[6] Zio Enrico Challenges in the vulnerability and risk analysis of critical infrastructures Reliability Engineering & System Safety, 152 (2016), pp. 137-150

[7] Ettore Bompard, Tao Huang, Yingjun Wu, Mihai Cremenescu Classification and trend analysis of threats origins to the security of power systems International Journal of Electrical Power & Energy Systems, 50 (2013), pp. 50–64.

[8] Chen-Ching Liu, Juhwan Jung, G. T. Heydt, V. Vittal, A. G. Phadke. The strategic power infrastructure defense (SPID) system - A conceptual design IEEE Control Syst Mag, 20 (2000), pp. 40–52.

[9] Ouyang Min Comparisons of purely topological model, betweenness based model and direct current power flow model to analyze power grid vulnerability Chaos, 23 (2013), pp. 293-296

[10] Cuffe Paul A Comparison of Malicious Interdiction Strategies Against Electrical Networks IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 7 (2017), pp. 205–217

[11] Colin R. Miller Electromagnetic pulse threats in 2010 Electromagnetic Pulse Threats in, 16 (2005), pp. 903-904

[12] Pu Cun-Lai, Cui Wei Vulnerability of complex networks under path-based attacks Physica A: Statistical Mechanics and its Applications, 419 (2015), pp. 622–629.

[13] Chee-Wooi Ten, Andrew Ginter, Rashiduzzaman Bulbul Cyber-Based Contingency Analysis IEEE Transactions on Power Systems, 31 (2016), pp. 3040-3050

[14] E. Bompard, R. Napoli, F. Xue Assessment of information impacts in power system security against malicious attacks in a general framework Reliability Engineering and System Safety, 94 (2009), pp. 1087-1094

[15] Yakup Koç, Martijn Warnier, Robert E. Kooij, Frances M. T. Brazier An entropy-based metric to quantify the robustness of power grids against cascading failures Safety Science, 59 (2013), pp. 126–134

[16] Networkx. Software for complex networks; 2002. [Online]. Available: http://networkx.github.io/; [accessed 20 August 2019].