Vulnerability Analysis on the Cloud Network Topology

Ke Chen, Hua Zhang

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China

cck232@163.com

Keywords: Hadoop, Vulnerability Analysis, Network Topology.

Abstract. In recent years, cloud computing is developping rapidly, and many companies put much effort to set up their own private cloud or public cloud. However, structures of cloud network topology are not of the same quality, and many network builders are tend to ignore the vulnerability of network topology which brings hidden danger to network security. Based on the minimum cut frequency vector method, this paper puts forward a method to measure the vulnerability index of structures of cloud computing network topology and to analyze the topology of the Hadoop cloud platform. Experimental results show that the bigger the index is, the more vulnerable the topology is. While improvement methods based on our standard can effectively reduce the above index, thus making the entire network more robust and reliable.

Introduction

Breakthrough is unceasingly made in technical reserves and in industry transformation of the cloud computing industry.More and more companies are engaged in research and development, manufacture, integration, and services of manufacturing related to cloud computing and the cloud computing industry chaingradually formed. Many companies are more inclined to store data on the cloud platform built by themselves, which involves lots of computing network building rather than store data on other companies' platforms , Powerful cloud computing platform is more than just the combination of computing nodes, it involves a complex network planning and building. From the perspective of planning, effective network topology can improve the disaster resistance capacity of network, and reduce the adverse impact of bad events.

Research on the vulnerability of network topology based on no center node began at an earlier time, which also tends to be mature, such as road traffic network topology[1][2][3] , P2P network[4], a complex network[5][6] topology vulnerability research, but the cloud computing network is often a center node network. After a thorough investigation and research in the industry, we found no standard is existed to measure the vulnerability of cloud computing network topology. No center node in the network topology based on vulnerability index calculation method does not consider the degree of node and failures probability of links In view of these two elements , we put forward a kind of standard which can be used to measure the vulnerability .of cloud computing network topology.

Our research object is the Hadoopcloud-computing network[7] If the network topology is fragile, attacks will lead to failure of connection of some equipment, and cause failure of communication between a large number of Slaver node and Master node, thus calculation ability is greatly reduced. Research based on this kind of problem is what we calledvulnerability researchonnetwork topology.

The second part of the paper mainly introduces concepts and definitions related to the topological vulnerability index, such as vulnerability vector index, which is the basis of vulnerability index calculation. The third part of the paper majorly gives an example to calculate the corresponding concepts and definitions. We compare data before and after improvement in the experiment to prove the feasibility of our vulnerability standards in the fourth part.The fifth part, at the end of the paper, lists the future work that need to be done.

Topological vulnerability theory foundation

In this section, we will state several concepts which will be used in the network topology analysis. In the first place is the minimum cut frequency vector brought forward by Bulteau and Rubino[8], which can be used to measure the vulnerability size between two points. According to the vulnerability between two points, combined with probabilistic edge importance index, finally we put forward an index standard to measure the entire network topology vulnerability. If the vulnerability indexof network topology obtained according to our standard[9] is too large, network robustness requirements cannot be met. Integrating the previous three concepts, we also put forward methods to improve the network topology vulnerability.

Vector Index. In the telecommunications field, Bulteau and Rubino[] put forward minimum cut frequency vector to represent the vulnerability index. In the network of their research, there is no central node and the degree of each node is ignored, namely vulnerability index between different nodes is evenly considered.However, the cloud-computing network is on the contrary. The Master nodes represent center node and each Slaver network has multiple host computer equipment. The number of the host computer equipment is the degree of each node. For instance, we define a network vulnerability index from some Slaver network to Master network:

TVI_{ms}= $(n_0, n_1, n_2, n_3,)$ (1)

$n_i$above means when the number of simultaneously disconnectedstripsis *i*, the Slaver network can't access the combination number of Master number. Given that we have just set up the Hadoop cloud computing network, where all the Slaver network can reach the Master network, so n_0 =0.

Edge Importance Index. According to definition (1), when edge *i* is removed simultaneously, some Slaver networks will not reach the Master network, and all compute nodes in those networks will not participate in the calculation, meaning that they can't help with the whole calculation. In case a single edge belongs to a set of such*i*edges, we analyze it from the probabilistic point of view, assuming represents the probability of simultaneous failure of *i*edges and with calculation for the number of nodes affected, the importance index of an edge can be obtained through the formula shown below:

$LCI=\sum_{i=1}^{t} p_i s_i$ (2)

wherein *t* represents the number of combinations where network disconnection occurs due to disconnection of*i*edges and represents the number of compute nodes affected by the simultaneous disconnection of*i*edges.

Vulnerability Index. Considering both the vector index and edge importance index, as well as the probability of simultaneous failure of *i*edges, we can find that the vulnerability index is negatively correlated to the size of *i*and the number of computing nodes affected will change with the failure of different *i*edges. Ultimately, through normalization processing, we can measure the vulnerability of the entire network topology with the following formula in a preliminary sens

$TVI= \sum_{k=1}^{n} \frac{1}{2^k} \left(\frac{\sum_{i=1}^{t} p_i s_i}{s} \right)$ (3)

Where in*k* is the number of edges failing simultaneously, *s* is the total number of computing nodes, p is the probability of simultaneous failure of *k* edge and represents the number of compute nodes affected by the simultaneous disconnection of*i* edges.

Improvement Method. According to the edge importance index of LCI, an edge with greater LCI is more important and an attack on such an edge will affect more compute nodes. There are two ways to address this issue: firstly, we can lower the importance index of some edges by adding more redundancy edges, which needs to be done according to the actual physical environment and the costs. Based on the greedy algorithm, we may lower the importance of the key edges by adding edges at positions in the model which are closest to such key edges; secondly, we can prevent the damage to the key edges by reinforcing such edges, here the key edges are corresponding to routers and switches in the actual physical environment.

By adding more redundancy edges, robustness of the networks previously connected by the key edges will be enhanced, so that the connectivity of the entire network can be maintained even when the original key edges have broken off. According to the topology vulnerability index(3), we can

calculate the declining quantity in network topology vulnerability index as follows:

$$\Delta TVI = \frac{1}{2^k}\left(\frac{\sum_{i=1}^{t} p_i s_i}{s}\right) (a \in D) \quad (4)$$

Calculation Example

Using Fig.1 as the example, we can calculate each index separately with the formula shown below, wherein the uppercase letters are network numbers, the figures represent the degrees of the networks and the lowercase letters represent the routers connecting two networks. Network A contains one NameNode and two DataNodes, adding to three nodes represented by the number 3 in the figure. Any other network contains only the DataNodes. As DataNodes are the nodes assuming computing tasks, their failure means that they will break away from the entire network and no longer be able to participate in the calculation. Generally, the NameNodes have high security coefficient, which make them less vulnerable to attacks. Therefore, we only need to analyze the routers connecting two computing networks, as multiple DataNodes may be unable to connect with the NameNodes once some routers are disconnected due to attacks.

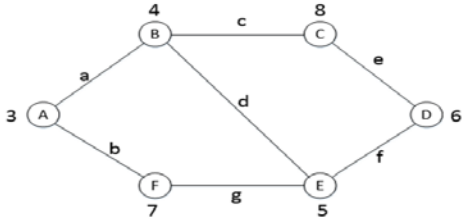

Table 1 VCI of all MS in trial network

Rank	MS	m_1	m_2	m_3	m_4
1	AD	0	4	5	0
2	AC	0	4	4	0
3	AE	0	2	5	0
4	AB	0	2	3	0
5	AF	0	2	3	0

Fig. 1 Trial network for calculation of TVI

Calculation for Vector Index of Vulnerability. By analyzing the reachability between Network A and B according to formula (1) we can find that: both Networks can be reached after removing any one edge, makingn_1=0. Two combinations (a, g), (a, b) can render both Networks unreachable when removing two edges at the same time, making n_2=2. Moreover, three combinations (a, c, d), (a, d, e), (a, d, f) can render both Networks unreachable when removing 3 edges at the same time, makingn_3=3. The Vector Index of Vulnerability between A and B: TVI_{AB}=(0,0,2,3,0,0). By conducting calculations with the same method on other networks, we have obtained the results shown in Table 1.

Calculation of Edge Importance Index. To calculate the important index of a specific edge, we need to first calculate the number of computing nodes affected by the disconnection of such an edge. For example, when edge (a, b) are both removed, all compute nodes of Network B, C, D, E and F will be affected as a result and the accumulative total of the degree of such nodes will be 30, meaning that 30 computing nodes will be affected when such two edges are removed simultaneously. With the same method we have made calculations for all combinations and obtained the results shown in Table 2.

Table 2 VCI of all links in trial network

remove	Probability p	nodes affected
(a, b)	$1/p^2$	30
(a, g)	$1/p^2$	23
(b, g)	$1/p^2$	7
(c, f)	$1/p^2$	14
(e, f)	$1/p^2$	6
(c, e)	$1/p^2$	8

We can calculate the edge significance index using formula (2) and with the calculation results in Table 2. For example, if we need to calculate the edge significance index of edge that is contained in combinations such as (a, b), (a, g), (a, c, d), (a, d, e) and (a, d, f), we can do so by using formula (2), assuming that p=0.5:

$$LCI_a = 0.25 * (30 + 23) + 0.125 * (4 + 12 + 18) = 17.5$$

The same is true for the calculation of the importance index of any other edge. As shown in Table 3, it is obvious that edge a has the largest importance index, meaning that this edge is the most critical edge.

Table 3 LCI of all links in trial network

Rank	edge	TVI
1	A	17.5
2	B	16.25
5	C	11.625
3	D	15.625
7	E	8.625
6	F	9.375
4	G	11.875

Calculation of Vulnerability Index.Where the probability for an edge to enter a failed state is 1/2, the probability for simultaneous failure of n edges shall be$1/2^n$ then by using formula (3) and taking into account the vertex degree we can calculate the vulnerability index of the entire network topology as follows: TVI = 0.22585.

Lowering the Vulnerability Index.There are mainly two ways to reduce the network's vulnerability index and enhance its reliability.The first is to strengthen a certain edge and reduce the probability of it being disconnected upon attacks, and the second is to add more edges. Judging from the angle of network topology, adding more edges that are redundant is the best solution. By such a method, we can recalculate TVI, i.e. the vulnerability index of the entire network topology and LCI, i.e. the important index of the original edge, and the extent by which such two indexes fall indicates the extent by which the reliability of the network is enhanced after the addition of more redundant edges. Assuming only one edge, for example edge h connecting Network A and E, is added, the vulnerability index of the entire network topology will be greatly reduced by the amount of ΔTVI=0.116 according to formula (4). In addition, the vulnerability index of network topology shall be: TVI = 0.10748; the importance index of edge a will be reduced to: = 10.875, despite that it is still larger than that of all other edges.

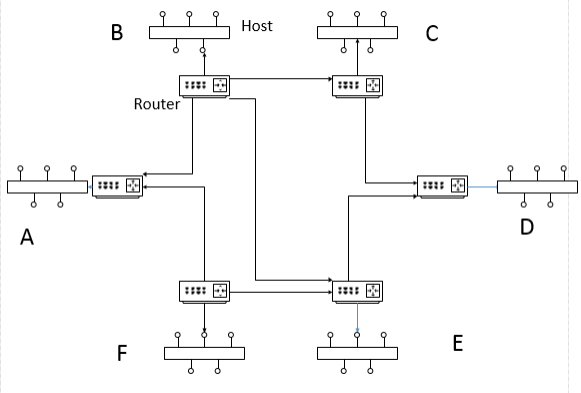

Fig. 2 The simulation diagram

Experiment

We have set up an appropriate simulation environment according to the topology diagram shown in Fig.2. Then we let some edges disconnect in a random manner and run the programs with the same workloads, so that we may record each run time for statistical analysis with the purpose of measuring the performance of the current Hadoop computing network. The simulation process is as follows. First, we set up the OpenStack cloud platform[10][11], for which we can create multiple virtual machines by using the IaaS services provided by OpenStack to build the Hadoop network. Then the size of the PI is obtained through the test program with the MapReduce program. Some routers are disconnected through the program (for example, we can simulate 1000 attacks on the network topology), and some compute nodes are rendered unreachable as a result. Lastly, we may conduct statistical analysis after recording each run time. According to topology diagram, there are 32

compute nodes, so the task is divided into 32 subtasks (each producing 100000 points).And statistical summary was conducted at the end; If edge set (a, b) fails, the workload of such 32 subtasks will be forced on the remaining two compute nodes, causing the computation time to be greatly increased.

Table 4 The number of nodes affected before/after adding the redundant edge

Failure	Run time before adding h	Run time after adding h
(a, b)	1274	240
(a, g)	925	237
(b, g)	314	220
(c, f)	498	194
(e, f)	291	230
(c, e)	320	237

Table 5. Program run time before/after adding h edge

	1000 attacks	nodes affectedbefore adding h	nodes affected after adding h
(a, b)	47	30	0
(a, g)	67	23	0
(b, g)	48	7	0
(c, f)	51	14	0
(e, f)	48	6	0
(c, e)	56	8	0

As can be seen from Table 4, 5 the original and improved topologies were subject to the same attacks. And in the case of improved topology, the computing performance of the network was not affected when some edge combinations were being attacked, meaning that the reliability of the network topology has been greatly enhanced.

Summary

This paper analyzes the vulnerability of the network from the angle of topology and evaluates topological vulnerability by using the TVI Index, so that network planners may compare the pros and cons of the different options. After integrating the simulation experiment which produced strong evidence proving that our theoretical determining method is highly consistent with the results of the experiment, the network administrators can, according to the edge importance index, strengthen critical edges in the existing networks and reduce the impact brought by damage to such edges as the greater the importance index of the edge is, the greater theloss its disconnection will cause and the extent by which the performance will fall. Further research in the future should be concentrated on more complex networks. We can add weight to the edges as an edge with greater weight is less likely to disconnect when being attacked, therefore we can analyze the topological vulnerability in a more accurate and comprehensive manner.

Acknowledgments

This work is supported by NSFC (Grant Nos. 61300181, 61202434), the Fundamental Research Funds for the Central Universities (Grant No. 2015RC23).

References

[1] Chen A., Yang, C.Kongsomsaksakul, S. and Lee, M, Networkbased Accessibility Measure for VulnerabilityAnalysis of Degradable Transportation Networks.,Networksand Spatial Economics, Vol.7, No.3: 241256, 2007.

[2] Tampere, C., Stada, J. and Immers, B, Methodology forIdentifying Vulnerable Sections in a National Road Network, Transportation Research Board of the NationalAcademies, Washington, D.C,

2007.

[3] Issacharoff, L., Lammer, S., Rosato, V. and Helbing, D, Critical Infrastructures Vulnerability: The HighwayNetworks, Springer Berlin/Heidelberg, 2008.

[4] Hu Z, Improved Algorithm of Unstructured P2P Network Topology StructureIntelligent Ubiquitous Computing and Education, 2009 International Symposium on. IEEE, 2009: 358361.

[5] Srivastava A, Morris T, Ernster T, et al, Modeling cyberphysical vulnerability of the smart grid with incomplete information, IEEE Transactions on, 2013, 4(1): 235-244.

[6] Mandal A, Xin Y, Baldine I, et al, Provisioning and evaluating multidomain networked clouds for Hadoopbasedapplications, Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on. IEEE, 2011: 690-697.

[7] Wang Y, Que X, Yu W, et al, Hadoop acceleration through network levitated merge, Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis. ACM, 2011: 57.

[8] Yingfei T, Chao Y, Xiaohong C, Methodology for evaluating and improving road network topology vulnerability, Intelligent Computation Technology and Automation (ICICTA), 2010 International Conference on. IEEE, 2010, 2: 664-669.

[9] Matisziw, T. C., Murray, A. T. and Grubesic, T. H, Exploring the vulnerability of network infrastructure todisruption, The Annals of Regional Science, Vol. 43, No. 2:307-321, 2009.

[10] Callegati F, Cerroni W, Contoli C, et al, Performance of Network Virtualization in cloud computing infrastructures: The OpenStack case Cloud Networking, 2014 IEEE 3rd International Conference on. IEEE, 2014: 132-137.

[11] Callegati F, Cerroni W, Contoli C, et al, Performance of Network Virtualization in cloud computing infrastructures: The OpenStack case Cloud Networking, 2014 IEEE 3rd International Conference on. IEEE, 2014: 132-137.