Novel Node Betweenness Based Fault Detection for Streetlamp Control Network

Runmin Wang

***Abstract*—In time of the Internet of things, some infrastructures of our city such as streetlamp control network is becoming more and more intelligent. It is necessary to apply some new technology to streetlamp network fault node detection which composed of computer network. In this paper, we combine network performance objectives in complex networks with real communication network's performance analysis. Experiments focus on the relationship between fault nodes and network performance such as cluster coefficient, inverse geodesic length and size of the largest connected subgraph. It is found that the node's betweenness plays a key role in the influence to above network performance objectives with a coarse linear relationship. We present the quantified analysis of influence generated in network performance and modify the traditional SNMP polling mechanism for streetlamp control network. The heuristic fault node location algorithm based on node's betweenness is proposed in this paper to promote fault location performance of traditional SNMP polling in the time of finding fault nodes in real networks of different types by algorithm parallelization in cloud computing platform.**

***Index Terms*—Streetlamp control network, network management, betweenness, fault detection**

I. Introduction

With the rapid increased scale of communication network, the network structure became more complicated. How to analyze and evaluate network's resilience after node failure have always been the hot research points in recent years [1]-[2]. Streetlamp network is one type of computer network which is classified as "scale-free network" for displaying the power-law distribution of degree by the research results in complex network [3]. Recent researches have also shown that network nodes which have large betweenness value closely related to fast information spreading in graph which is helpful for fast fault node location [4]–[5]. The traditional fault node location theories are based on thesis that computer network is random network, it's necessary to research the relationship of resilience after node failure and the latest discovered rules of complex network for fast fault node location.

II. Related Work

A. Network Resilience

The research of vulnerability and robustness in communication network after node failure or network attack is important to network resilience analysis and fast fault node location. In 2002, Holme [2] studied the response of complex networks subject to attacks on vertices and edges with four different attacking strategies. They found that the removals by the recalculated degrees and betweenness centralities are often more harmful than the attack strategies based on degrees and betweenness centralities of the initial network. But they only use one real communication network and did not proposed some application points. M. E. J. Newman [3] studied that the failure of single node in the network can not cause bicomponents to be disconnected. The dataset in researches above lack of real communication networks and they proposed few applications in fault location.

Manuscript received February 27, 2012; revised April 29, 2012.

R. Wang is from College of Computer Science and Technology, Chongqing University of Posts and Telecommunications(Tel.: + 86-023-62460934; e-mail: 799862370@qq.com)

B. Fault Node Location Algorithm

In recent years, researchers proposed some new fault location algorithm based on graph theories. In 2007 Cui [6] proposed a self-adaptive SNMP polling policy for scale-free communication networks. Zhang [7] proposed a fault monitoring algorithm for hierarchical network. But these new fault location algorithms have their drawbacks. The summary of node fault detection algorithm can be found in Fig.1 composed of three main types which are AI technologies based algorithms, Model traversing technologies based algorithms and fault propagation technologies based algorithms. Therefore, we need to construct some new fault node location algorithm based on the new founding of complex network theory to promote fault location efficiency.

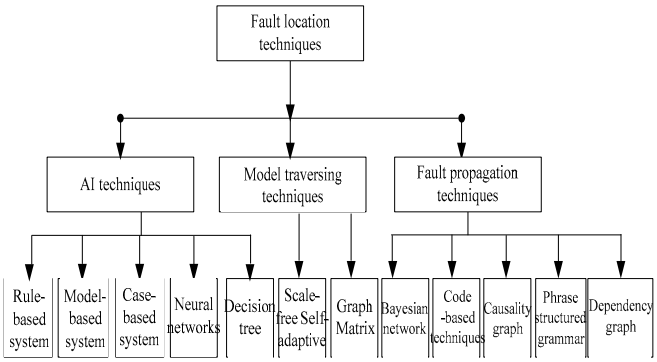

Fig. 1. Node fault detection algorithm summary

C. Node Betweenness Algorithm

Aiming to decrease the high time complexity of existing node betweenness on large scale graph, there have been some new algorithms (e.g. approximation algorithm) to handle high time complexity. In 2004, David Eppstein and Joseph Wang [8] proposed the estimation method of Closeness Centrality by fast random sampling. In 2007,

Brandes [9] proposed the approximation estimation method based on foundation of the idea which there is a bound on the deviation of the average of a given number of bounded random variables from its expectation. Brandes used some pivots to estimate the betweenness of target node.

But those mentioned above methods does not construct the approximated whole graph node betweenness algorithm in large scale graph based on Map Reduce. And running time cost of them is always high result in hard to application. MPI based programming method is often difficult for ordinary program developers to get used to the parallel programming mechanism. To the contrary, Map Reduce is much easier to learn and to use and it is necessary to construct Map Reduce based whole graph node betweenness approximated computation algorithms to promote computation efficiency.

III. Experiments on Communication Network' Resilience

A. Definitions of Network Structure Quantities

The network models discussed in this paper are undirected and unweighted one which can be expressed as $G=(V,\varepsilon)$. V is the set of nodes with number $N=|V|$ and ε is the set of edges with number $L=|\varepsilon|$.

1) Average Inverse Geodesic Length

Average inverse geodesic length l is the important network structure quantities after node failure.

$$l \equiv < d(v,w) > \equiv \frac{1}{N(N-1)} \sum_{v \in V} \sum_{w \neq v \in V} d(v,w) \tag{1}$$

$d(v,w)$ stands for the geodesic path length between different node v and w.In small-world network l is around 6 and in www web it is around 17.After some nodes are removed, if there is no path between node v and w, $d(v,w)$ reaches $+\infty$.There is another length quantity l^{-1} instead of l:

$$l^{-1} \equiv < \frac{1}{d(v,w)} > \equiv \frac{1}{N(N-1)} \sum_{v \in V} \sum_{w \neq v \in V} \frac{1}{d(v,w)} \tag{2}$$

where there is no path from v to w, $1/d(v,w)$ equals zero.

2) Relative Size of Largest Connected Sub Graph

Size of largest connected sub graph refers to the node number of largest connected sub graph in the whole graph and it is the important quantity to reveal inner connectivity characteristics of graph. If the connected sub graphs set are $\{g_1, g_2, g_3, g_4, \ldots\ldots g_n\}$ in $G=(V,\varepsilon)$, then the relative size of largest connected sub graph can be defined as:

$$S = \frac{\mathrm{Max}\{|V_{g_1}|,|V_{g_2}|,|V_{g_3}|,\ldots\ldots,|V_{g_n}|\}}{|V|} \tag{3}$$

3) Network Average Cluster Coefficient

The network cluster coefficient reflects the macroscopically cluster characteristics of whole network to reveal the density of links among neighbor nodes. The cluster coefficient of node can be defined as:

$$\mathrm{C_i} = \frac{2E_i}{k_i(k_i-1)} \tag{4}$$

k_i stands for the neighbor node number of node i and E_i stands for the number of existing links among neighbor nodes of node i.Network average cluster coefficient can be defined as:

$$\mathrm{C}_{G=(V,\varepsilon)} \equiv \frac{1}{N} \sum_{v_i \in V, i=1}^{N} \frac{2E_i}{k_i(k_i-1)} \tag{5}$$

B. Simulated Failure Generation Mechanism

The four node attack strategies [10] are as follows:

1) ID removal: attack starting from the node with the highest degree and node attack strategy uses the initial node degree distribution.

2) IB removal: attack starting from the node with the highest betweenness and node attack strategy uses the initial node betweenness distribution.

3) RD removal: using the recalculated node degree distribution at every removal step.

4) RB removal: using the recalculated node betweenness at every step.

C. Experiment Result

1) Dataset

TABLE I: Network Dataset

	PFP(a)	CN07(b)	ITDK0304(c)
Number of node N	500	135	9204
Number of link L	1500	138	28959

Selection of dataset is important to experiment results and we selected more real network datasets than which adopted by Albert and Holme. The network dataset (in Table I) includes the PFP (positive feedback preferential) which reflect the "rich-club" phenomenon of Internet AS level network (a), the AS topology of China in 2007 from Skitter(b), the AS topology of Global Internet in 2003 (c).

2) Experiment Result

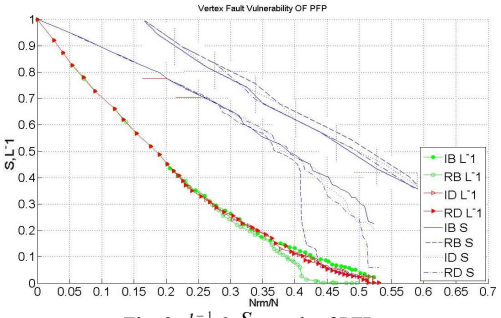

Fig. 2. l^{-1} & S result of PFP

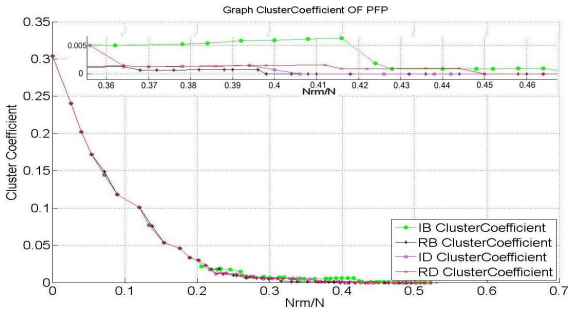

Fig. 3. $C_{G=(V,\varepsilon)}$ result of PFP

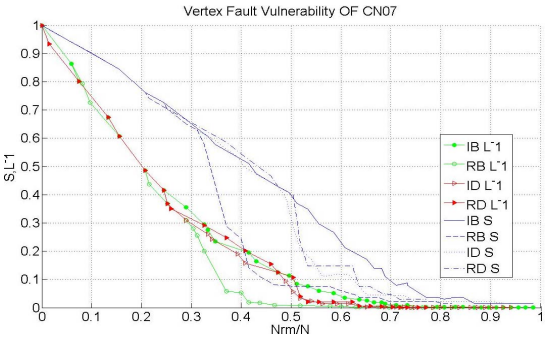

Fig. 4. l^{-1} & S result of CN07

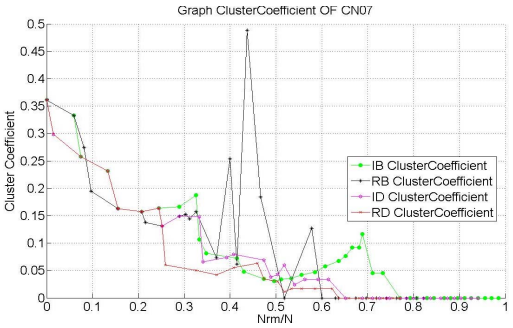

Fig. 5. $C_{G=(V,\varepsilon)}$ result of CN07

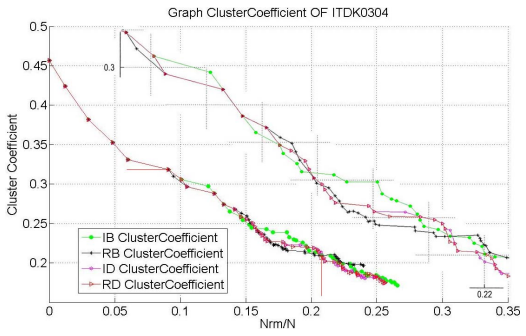

Fig. 6. $C_{G=(V,\varepsilon)}$ result of ITDK

The x-axis in Fig. 2 to 6 stands for the removed node ration (Nrm: the number of removed nodes) to initial graph. The y-axis stands for the relative value of S and l^{-1} (the ration of S and l^{-1} after every step of node removal to initial S and l^{-1} in the network).

D. Experiment Result Analysis

We can summary that from Fig. 2 to 6:

1) l^{-1}: RB is the most harmful strategy when l^{-1} is bigger ($0.3< l^{-1} <1.0$), RB>IB>RD>ID.When l^{-1} gets smaller ($0.0<l^{-1}<0.3$), RB>RD>ID>IB.

2) S: RB is the most harmful strategy. Based on Figure 1 and 3, it can be concluded that when S is bigger (S >0.7) the four strategies are almost the same harmful. When 0.4< S <0.7, RB>IB>RD>ID. When 0.0< S <0.4, RB>RD>ID>IB. The S curve generated by IB almost demonstrates the linear relationship.

3) $C_{G=(V,\varepsilon)}$: Among the four $C_{G=(V,\varepsilon)}$ curves, the sudden rises caused by IB are much more than the rises caused by other strategies. The influence to $C_{G=(V,\varepsilon)}$ approximately shows that IB>RB>ID>RD. In a summary of influence to l^{-1}, S and $C_{G=(V,\varepsilon)}$, RB and IB are more harmful.

4) For fault node location: for the reason that node betweenness based strategies are more harmful, the node with larger betweenness in SNMP Polling process should be visited earlier to avoid more harmful destroy to whole network.

IV. Betweenness Based Fast Fault Node Location Algorithm

A. Algorithm Structure and Framework

From the experiment results above, we attained some valuable information for fast fault location in network management: 1) in communication network, the node with high betweenness always carries high load of data flow and the rate of node breakdown rises up. SNMP Manager should pay more attention to these nodes and nodes with high betweenness should be arranged in the front of polling order. 2) When the data flow rate of node changed in the network in case of no fault nodes occurred, we need to adjust the polling interval time to get more accurate data of data flow parameters. In order to calculate the polling order of one node, we adopt W_i in formula (6) to stands the importance weight of node i.

$$W_i = \frac{(B_{1i}\times\alpha + B_{2i}\times(1-\alpha))\times(\Delta ifInOctets \times\lambda + \Delta ifOutOctets\times(1-\lambda))}{CurrentTimer[i]} \quad (6)$$

In formula (6), B_{1i} stands for betweenness of node i before node failure happen in network and B_{2i} stands for the betweenness of node i. $\Delta ifInOctets$ and $\Delta ifInOctets$ are the variations of inflow and outflow data in two continues polling intervals. α And λ are adjusting parameters. Function of parameter λ is almost the same to $\Delta ifInOctets$.

B. Result of Fault Location Experiment

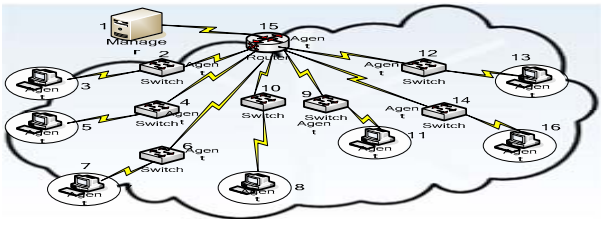

Fig. 7. Star topology evaluation environment

TABLE II: FAULT LOCATION RESULT OF FIG. 7

Failure time point	Node 4	Node 7	Node 9	Node 13	Node 16
SNMP Polling report time	5s	10s	20s	30s	40s
Time used	11.951s	13.214s	11.128s	13.772s	16.420s
IB location report time	13.842s	19.125s	27.210s	37.623s	49.124s
Time used	8.842s	9.125s	7.210s	7.623s	9.124s

V. PARALLEL BETWEENNESS APPROXIMATION ALGORITHM

A. Algorithm Framework

For the reason that node Betweenness is vital for some important node in commonication network,it is necessary to parallelize betweenness computation algorithm in whole graph and the parallelized betweenness process are followed in Fig. 8.

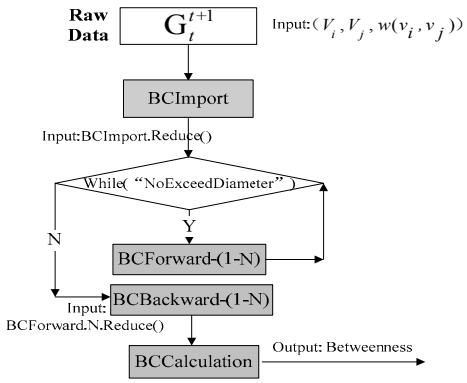

Fig. 8. Betweenness parallel details

VI. CONCLUSION

In this paper, we compared the variation trend of some key performance quantities of scale-free communication network's resilience of streetlamp control network by four node failure generation strategies. Experiments results showed that node betweenness based strategies such as RB and IB are more harmful. And then we proposed a new betweenness based SNMP fast fault node location algorithm which has been evaluated to promote network management efficiency. Furthermore we implement the cloud computing based betweenness algorithm for large scale network.

REFERENCES

[1] M. Subramanian, "Network Management-principles and practices," *High Education Press 2002*, pp. 562-563.2002.
[2] P. Holme, B. J. Kim, and C. N. Yoon, *Attack vulnerability of complex networks. Physical Review E*, 2002.
[3] M. E. J. Newman, *Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E*, 2006.
[4] D. J. Watts and S. H. Strogatz, Nature, pp. 440-442, 1998.
[5] A. A. Nanavati and R. Singh, *IEEE T. Knowl. Data En,* pp.703-718. 2008.
[6] J. Q. Cui, Y. X. He, and L. B. Wu. *J. Wuhan Univ. Technol*, vol. 53, no.3, pp. 293-296. 2007.
[7] G. Q. Zhang, *Journal of Computer Research and Development*, vol. 43, no. 10, pp. 1790-1796, 2006.
[8] E. David and J. Wang. *Journal of Graph Algorithms and Applications*, vol. 8, no. 1, pp.39-45, 2004.
[9] Ulrik Brandes, *Journal of Mathematical Sociology,* vol. 25, no. 2, pp. 163-177, 2001,
[10] R. Albert, H. Jeong, and A. L. Barabási, *Nature*, pp. 378-382. 2000.