Topological Optimization of Hubs in P2P Network

Liu Yanmei[1], Cui Huanhuan[1]
1.Chutian College Huazhong Agricultural University
Wuhan, China
61704088@qq.com

Chen Yuda[2]
2.Wuhan center of Geological Survey
Wuhan, China
Chyd555@163.com

***Abstract*—This paper further studied the hubs in P2P network, and presented a strategy based on hierarchical P2P network by controlling the logical topology structure and through enhanced mechanism to avoid the formation of hubs. It firstly introduced the idea of control model, and the specific implementation of control model .Then carried out complexity analysis of the algorithm. Finally, it validated the via simulations ,The simulation resulted that the control model of topology optimization could effectively control the formation of hubs, and improve the ability of P2P network anti-coordinated attack, thus effectively enhance network robustness and load balancing.**

Keywords- P2P Network; TTL-k; Topology Optimization; Hub; Hierarchy;

I. INTRODUCTION

At present, the P2P technology has something to be improved, since there does not have a recognized model that can Compatible with the characteristics of P2P network, such as fault tolerance, self-adaptability, expansibility, security and anonymity. However, with the development of P2P protocol and P2P software, the P2P software has been very popular nowadays.

P2P network is an overlay network based on Internet, P2P mode is different from the traditional mode of client/server, it has changed the traditional method of centralized storage and processing resources, effectively organize the boundary network resources, directly exchange information between resource providers and receivers based on the P2P network.

In the ideal P2P network also referred as a peer to peer network, nodes are both client and server, and both provide resources and consume resources, but it has not a central server. P2P communication mode provides a fast and convenient search and download files function for normal users, which promotes the rapid development of P2P, but also it can pose some problems: a few of P2P network nodes maintain a very high number of links [1]~[7], these nodes overheating which we call hub.

The existence of hubs in P2P network will lead to weaken the anti-attack ability of the whole system and make the system extremely fragile [2]. Therefore, if this phenomenon is uncontrolled, the number of hubs will be more and more, as mentioned above, each node is equal, no obligation to provide sharing or downloading of file for other nodes in a long time, When these hubs removed or malfunctioned could lead to the P2P network will be divided, and make P2P network service of local even more large paralysis [6]. Therefore, how to avoid the emergence of hubs becomes one of the key problems to be solved in P2P network.

The reference [8] posed through the logical topology of P2P network to control hubs in the network, the simulation results show that it can effectively enhance the anti-vulnerabilities of network, but it did not fully consider the heterogeneity of those spare nodes. In this paper Topology model which can control the formation of hub was improved, considering the heterogeneity of spare nodes, and introducing the reinforcement mechanism, by looking for a suitable node to backup the overheating resources of the source node, and it could have the effect of load balancing, and to improve the system resource utilization ratio.

II. RESEARCH STATUS

With the rapid development of P2P applications, which exposed more and more problems, hubs is one of the more serious and urgent problem. The degree of node in the network, the number of connections that a node has, there is a similar phenomenon to its power-law distribution [4].

The node degree obeys a power-law distribution which is the relationship between the number of nodes with a degree of K and the K approximation satisfies the power function $P(k) = k^{-\lambda}$, λ is a constant factor depends on the network itself, which means that the number of connections that most nodes in the network is very few, but a few of nodes have a lot of connections, the nodes that have a lot of connections are called hub.

Although a series of unstructured P2P networks such as Gnutella and Freenet do not meet the strict power-law model, but it can be seen as the composite of other models, which have a characteristic of power-law model on the whole, for example, the high fault tolerance in the face of random node failure and so on.

In the hierarchical processing for hub, the reference [9] presented dealing with hierarchical processing and distributed processing approach for hubs in scale-free network, thus optimizing the scale-free network topology, and enhancing the anti-fragility of the network. The reference [8] presented that when a node reached a limit, two spare nodes would be selected from the network, and then compared the two distances from the requesting node Q, got the small distance of node as a standby node, in order to achieve system optimization.

III. IMPROVED CONTROL MODEL OF HUBS IN P2P NETWORK

Presently, it is common that Controlling the behavior of a single node using incentive mechanisms [10] ~ [12], although it can effectively inhibit the free-rider behavior, but the part of free-rider nodes exit the system as a direct result, resulting a drop in the number of nodes, which is contrary to expand user base, and to enhance the influence of commercial purposes.

A. *Existing Problems and Improvement of Control Model*

Conventional topology control model [8, 13-15] has the following problems:

1) When there is a big gap between the requesting resources node Q and two spare nodes, this choice of optimization measures can play a role in the system, but it is from the source node query information so as to increase the P2P network query traffic expense; when the distance between the two spare nodes and the requested resource node Q is not obvious, the optimization measures effect is not obvious, and the node Q is the preferred choice node from its neighboring nodes as a backup node.

2) Take the absolute value of appropriate fields by subtracting IP address with another in the same way and the results compared to the node can't effectively judge the distance.

3) Does not consider the location of the standby node problem: the source node in the use of TTL-k search qualifying spare node has its inbuilt limitations, may be in the k-hop cannot find such qualified nodes. In addition, the value of k should not be too large, because the network overhead flooding method is increased with the increase of k increased exponentially.

According to (1) and (2) in this paper, using the single tree branches of logic control structure, logic control instead of using binary tree structure, which is just need to find a standby node, and let this node and the node Q to establish a connection.

In addition, according to the (3) if cannot find overload resource nodes of the source node in k jump, then stop looking for such a node, and looking for a surplus capacity in k jump range (including bandwidth, CPU, the response delay, etc.) the strongest node as a backup node, in order to offer the source node copies corresponding resources to the backup node, which can play a load balancing role.

B. *Improved Control Model Thought*

1)When finding a node Vi in the network will become the distribution node, to find the resources which have the maximum number of requested connections in the node Vi shared resources, immediately from the node Vi, to search other nodes in the network have this part of the resources, to pick out one from these nodes to support the maximum connection number of nodes as a standby node, and record the IP address of the node, so the source node and standby node to form a single logical structure tree branch , but the original quasi hub is the root node. Then this part of the new connection request is forwarded to an alternate resource node and the standby node in response to a request directly to the source node to establish a connection.

2）If cannot find those nodes which can support larger number of nodes connected in k jump, you stop looking for the kind of node. Instead looking for a strongest remaining capacity node as a backup node in k jump range for the source node will copy the appropriate resources to the backup node. Then the new resource request which will connect this part of resources is forwarded to the backup node, and in response to a request directly to the source node to establish a connection request.

3）If the standby node or backup node has distributed phenomenon, then the node adopts 1) above manner, 1) cannot handle and the method for processing in 2) can adopt, if the phenomenon still occurs for the resulting standby node or backup node, the method continues to repeat the above processing, the tree branch will eventually form a single logical structure shown in Figure 1, Where the solid line represents the phenomenon of impending distribution node and the standby node connections, dashed line indicates an impending distribution phenomena connected nodes and the backup node.

After treatment, if the distributed phenomenon of the node Vi has not been significantly changed, then their other shared resources above process again, repeat this process until Vi distributed phenomenon disappears. Finally, the whole system will become more hybrid structure which a lot of single branch tree cross together (Figure 2).

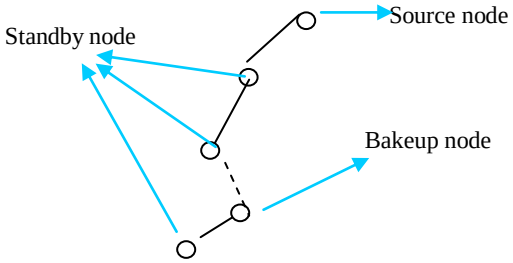

Figure 1. The single tree hierarchy

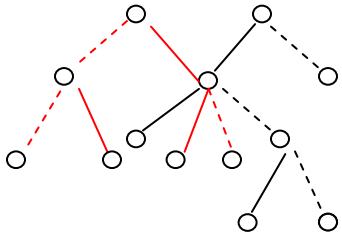

Figure 2. The hybrid single tree hierarchy（The dotted line represents the standby node is not found using the backup node）

C. *Control Model Thought*

The number of support connections L(I) said, how many connections can be supported except the existing connections by the node. L(I)= R(I)-D(I), the R(I) represents of the maximum number of connections, namely threshold, D (I) represents the degrees of the node.

Specific control model is as follows:

Assuming the threshold of a node Vi in a P2P network is R (I).At some time t, node Vi receives a new connection request Q, this moment, the node Vi has shared the m ($m \geq 1$)resource S, set Q is the request of resource S $\in$ S (S), this time the degree of node Vi is D (I).

1. If D(i)<R(i)-1, directly accept the request.

2. If D(i)=R(i)-1, and the node Vi has established standby node (or backup node) V about the resource s, then the node Vi will transfer the request Q directly to standby node (or backup node), let the VQ and V establish connection directly.

3. if D(i)=R(i)-1,and the node Vi hasn't establish standby node and backup node on the resources S, the processing procedure is as follows: the node Vi will search the resource S(1) which has the maximum number of nodes for it:

1)By searching the scope of resources S(1) which has supported the maximum number of connections to the node Vi to find l(l>0) nodes, if the node Vy which has the maximum number of connections(and not 0) can be found from those nodes, Vy will be the standby node about the resources S(1) of the node Vi, and form the logical structure of single branch with the node Vi, the node Vi is the root node, and the node Vy is the child node, the IP address of the node Vy is IP(Vy), it will terminate a latest connections of the node Vi about the resources S(1), Vi respond to the request Q.

2）Through within the scope of the nearby can't find any node that occupy the resources and support a certain number of connections, then from the node Vi to find N nodes that have strong ability (computer processing power, bandwidth, etc.) in the neighboring network, to choose a node which is the most strongest ability node as a backup node Vf for the node Vi, copies the requested resource S of request Q to the node Vf, then let the node Vf to establish a connection directly with VQ. If the standby node or the backup node number of connections to each smaller than the threshold 1 R (I) - 1, then repeat this process for the node.

This control model has made logic structured between standby node(or backup node)and the source node. If not properly structured control, each node and the source node will form a random graph structure between nodes (as shown in figure 3), after using a structured control will form a tree structure (as shown in figure 4), and the structure of the graph search algorithm's time is more complex than the tree structure's. On the other hand, the P2P network is dynamic, standby node (or backup node) may quit at any time, and the existing mature algorithms can be used for the binary tree node updates, it is easy to maintain, therefore, the standby node (or backup node) and the source node to establish logical tree structure is more appropriate. In figure 4 node B is a standby node for node A, the node C is the backup node of node B.

D. Hierarchical Proceeding for Nodes

The source node Vi and the standby node Vy or the backup node Vf constitute a hierarchical relationships, the node Vi is the root node, the node Vy or the node Vf is the child node.

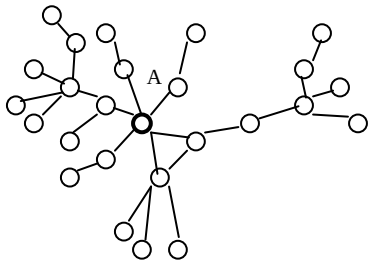

Figure 3. A tree structure not constructed by structured control

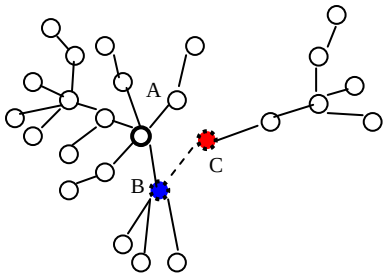

Figure 4. A tree structure constructed by structured control

Set the record format of standby node Vy and the backup node Vf for one node on a certain resource S is a triples respectively:

(S, IP (Vy), IP (Vf)), for a certain resource S each hub maintain a spare backup node list List_children, it is be used to record the set of triples for the node, the initial value of the IP (Vy) and IP (Vf) is null. Each standby node and backup node maintain a table List_parent_BY and List_parent_BF, which is used to record the parent node Vi about resource S.

TABLE I. LIST_CHILDREN

Resource	Standby node	Backup node
s	IP(Vy)	IP(Vf)

TABLE II. LIST_PARENT_BY

Resource	Father node
s	IP(Vi)

TABLE III. LIST_PARENT_BF

Resource	Father node
s	IP(Vi)

As mentioned in the section 'C.', when D(I) = R(I)- 1, need to search the table List_children, find the item S, the failure to find returns false. Find the success, continue to view the IP (Vy) is whether or not null, if the value isn't null then returned the IP (Vy), if it is empty then continues to check the next IP (Vf) is whether or not null, if IP (Vf) is not null, it returns the IP (Vf). If it is empty, then the situation using the section 'C' strategies to find a standby node or a backup node.

When the node Vi's child node about resource S' is be used as the standby node, the table List_children (Vi) that is maintained by the node need to add a new record (S', IP (Vy), null), at the same time, Vy need to maintain the table

List_parent_BY (Vy), and add records (S', IP (Vi)).To add records to the backup node can use the similar methods.

When the node Vj as other node's standby node in binary tree exit system, need to check the resource of List_children (Vj) table about the List_parent_BY (Vj) table and List_parent_BF (Vj) table of the node Vj. If the query succeeds, the query results IP (Vy ') and IP (Vf) are be sent to the corresponding parent node, the parent node will update its IP(Vj) and IP (Vf) information for IP (Vy ') or IP (Vf) in List_children; Otherwise send Null to the parent node of the corresponding resources, the parent node update null to IP(Vj) in its List_children_BY . To the exit of the backup node can use the similar methods.

The most complex place in this algorithm is that when to the node Vi, D (i) = R (i) -1 and Vi hasn't established the standby node or the backup node about the resource S, then you need to search List_children table, set the table at most having m items, time complexity is O (m) in the worst case. Secondly, to find the maximum number of connections about resources S (1) time complexity is O (m). Again, in the searching network, the time complexity T depends on the complexity of the existing P2P search algorithm, one node is selected form the k nodes that can support the connection number, the comparison of the number is k-1. If there hasn't the node that support the connection number, this process should be repeated m times, its time complexity is m * (T + k-1). Therefore, the overall time complexity of the algorithm is m + m + m * (T +2 k-1). From the meaning of m and k, we can see m and k are countable, so the total time complexity is O (T) that depends on P2P search algorithm time complexity.

IV. SIMULATION RESULTS AND ITS ANALYSIS

Simulation based on BA algorithm to generate network topology, the whole structure including 8000 nodes, assuming that there are 1000 files randomly distributed on the 8000 nodes, each different from each other sharing 100 files, the initial network is a complete network structure that is established by eight nodes, then each a new node connect 5 edges. The threshold of each node has a random value, and assumes that the threshold value is 30.The parameters related to the experiment and the default values can be seen in table 4:

TABLE IV. EXPERIMENTAL ENVIRONMENT PARAMETERS AND THEIR DEFAULT VALUES

Parameters	The number of nodes	Routing algorithm	The number of the total files	The number of every node owns	The threshold of nodes
Value	8000	Flooding	1000	100	30

The experiment using Gnutella simulation network model and Flooding search algorithm, using the control model of this paper and not using this model in both cases, respectively simulate the connection request 8000 times in the network resources, the following are two cases of node degree distribution results comparison chart. For comparison, the node degree is divided to two intervals (D≤20 and D>20).

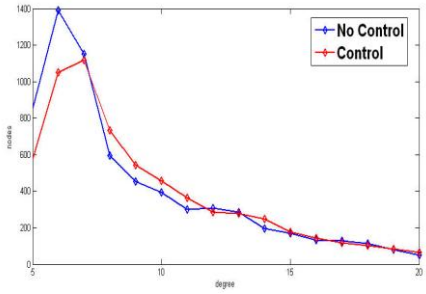

Figure 5. Degree D≤20 interval

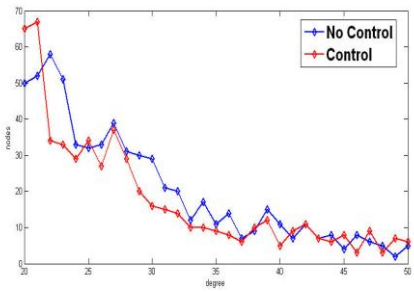

Figure 6. Degree D>20 interval

Figure 5 shows a comparison of the interval those less than 20 degrees in the two cases in Fig. In the area of not more than 7 degrees, the number of nodes under the use of this control model was significantly lower than the control is not to take the case, and in the subsequent regional 8-11, the number of nodes under using this control model is slightly higher than the situation without taking control.

Figure 6 shows a comparison of degrees higher than the section 20, the portion from 20 to 29 degrees is no large difference between the results of the two cases from the figure, but in 30 to 37 sections, comparison of the two cases is very obvious. As the simulation node set threshold value of 30, so around 30 in the vicinity has formed a clear dividing line. In the section of higher than 37, the contrast of two kinds of cases is not obvious, however, the highest degree after taking control is lower than the case without control (The highest degrees is about 373 after taking control in the network, the highest degree can reach 468 without the control).

In figure 5 and figure 6 to present the above phenomenon is be formed in the processing of controlling hubs, the connect query to hub is forwarded to the lower degree standby node, to change the original logical topology structure, so that the degree of nodes in the network is equalized. From the figure, the number of nodes is increased in a small degree of area, the number of nodes in the high number of area will reduce and the maximum degree of the network is reduced.

From the simulation result shows that the model can transfer the connection from the high degree node to the

lower degree node, changing the logical topology of the P2P network, thereby reducing the high connections of nodes in the network, and also reducing the maximum degree of the network, indicating that the number of hub has been reducing in the network, distribution phenomenon has been effectively controlled. Further, reducing hubs, the capacity of anti-coordinated attack in the network will be enhanced; thereby improving the robustness of the network, and changing the vulnerability of P2P network due to the existence of hubs. Therefore, the simulation results show that this control model is effective.

V. CONCLUSION

The existence of hub in the P2P network can reduce the coordinated attack resistance of the whole system and make the system vulnerable, affect the quality and service performance of P2P system. This paper puts forward an improved by controlling the logic of the P2P network topology based on hierarchical structure to avoid network hub forming method. The simulation results show that this control model can effectively control the formation of hub in P2P network, which can improve the resistance of the whole network coordinated attack ability, enhance the robustness of the network, and can have the effect of load balance, improve the utilization rate of the whole system and ensuring the healthy development of the P2P network.

ACKNOWLEDGMENT

Yi Wang, YuLing Li and HuiTing Wu are thanked for helpful review comments and to YuDa Chen for editorial handling and thoughtful suggestions, which have helped to improve this article. The research is supported by Hubei Provincial Department of Education (Grant No. 2013455, No.B2014268), Chutian College Huazhong Agricultural University (Grant No. 201303, and K201304).

REFERENCES

[1] A. Klemm, C. Lindemann, M. Vernon, and O. Waldhorst, Characterizing the Query Behavior in Peer-to-Peer File Sharing Systems, Proc. ACM Internet Measurement Conference (IMC), Taormina, Italy, Oct 2004: 55-67.

[2] Wenjie Wang, Hyunseok Chang, Amgad Zeitoun etal. "Characterizing Guarded Hosts in Peer-to-Peer File Sharing Systems", IEEE Global Communications Conference (Globecom 2004): 1539-1543.

[3] Adar, E., Huberman, B.,"Free Riding on Gnutella". First Monday, October 2000.

[4] Matei Ripeanu, Ian Foster and Adriana Iamnitchi. Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design. IEEE Internet Computing, vol. 6(1) 2002.

[5] Sen S, Wang Jia. Analyzing peer-to-peer traffic across large networks. IEEE/ACM Trans. on Networking, 2004,12(2):219-232.

[6] D. Hughes, G. Coulson, and J. Walkerdine. Free Riding on Gnutella. Revisited: The Bell Tolls IEEE Distributed Systems Online, 6(6),. 2005.

[7] Ramayya Krishnan, Michael D Smith, Zhulei Tang et al. The Impact of Free-Riding on Peer-to-Peer Networks. Proceedings of the 36th Annual Hawaii International Conference on System Sciences(HICSS-37 2004): 199-208.

[8] Chun Yang, Yuhua Liu, Kaihua Xu, Hongcai Chen.Model of controlling the Hubs in P2P Network,Computer Science, 2009 ,36(2):62-65.

[9] Shaohua Tao, Yuhua Liu, Kaihua Xu, Demao Tan. The Strategies against Vulnerability of Hubs in Complex Networks, Computer Engineering and Applications, 2007,43(2):151-153.

[10] Richard T.B. Ma, C.M. Lee, John C.S. Lui, David. K.Y. Yau. Incentive P2P Networks: A Protocol to Encourage Information Sharing and Contribution. Performance Evaluation Review, 31(2), September, 2003: 23-25.

[11] Emmanuelle Anceaume, Maria Gradinariu, Aina Ravoaja. Incentive for P2P Fair Resource Sharing. Proceedings of the IEEE International Conference on Peer-to-Peer Computing (P2P'05): 253-260.

[12] Haigang Gong, Ming Liu, Yingchi Mao et al. Research Advances in Key Technology of P2P-Based Media Streaming. Computer Research and Development, 2005, 42(12): 2033-2040.

[13] Hao Rao, Chun Yang, Shaohua Tao. Distributed controlling model for hubs in P 2 P networks.Application Research of Comouters,2009,26(12):4686-4689

[14] Lilong Chen. Reserarch on the Node Load Balance Control Mechanism in Unstructured P2P Networks. Wuhan: Huazhong Normal University.2011.

[15] Miaomiao Yin, Shiping Chen.Generic Topology Matching Method Based on Structured P 2 P Network.Computer Systems & Applications,2012(4):135-138.