Reliable Resource Search in Scale Free Peer-to-Peer Network

Wei Song[1], Wenbin Hu[1], Zhengbing Hu[2], Xi Zeng[1]
1 Computer School, Wuhan University, Wuhan, China
2 Department of Information Technology, Huazhong Normal University, Wuhan, China
Email: songwei@whu.edu.cn, hwb77129@126.com, hzb@mail.hust.edu.cn,zx0071@163.com

Abstract—**The resource distribution and peer links in P2P network have an obvious scale free character follows power law distribution. Using this inherent character of P2P network to design resource search strategy is great significant for improving the search efficiency and reducing the costs. We analyze the scale free character in P2P network, and propose the reliable random walk search algorithm which can achieve high and reliable resource search through transferring query messages based on P2P power law distribution. We design simulation experiments to evaluate the performance of our reliable random walk. The simulation experimental results show that the reliable random walk based on power law distribution is a scalable resource search algorithm with high and reliable search efficiency and low search costs.**

Index Terms—**Peer-to-Peer, resource search, complex network, scale free, power law distribution**

I. Introduction

With the rapid development of computer and network technology, computing model is in progress of significant alteration. As a new network service model, P2P (peer-to-peer) network changes the way people share resources over Internet. Each node is not only a client, but also a server in P2P network system, whose resources are not stored on centralized servers, but on every distributed node local. In an unstructured P2P networks, nodes often use flooding and various optimized strategies to search desired resources. How to search resource effectively and efficiently in the case of lower network cost is a key issue in the study of P2P network.

A large number of complex systems in nature, such as the Internet, transport system, the spread of disease, various social networks and so on, can be described by network. The earliest research on network is made by mathematician, whose basic theory is Graph Theory. Classical graph theory tends to use regular topology to model real network, resulted in the emergence of regular network model. In the mid-twentieth century, Erdos and Renyi proposed that the establishment of a network connection is random and unordered. Based on this standpoint, the so-called random network model [1] is established, which had been the theoretical basis of scientists' study of real network for a long time. Until recent years, researchers discovered that large numbers of real networks are neither regular networks, nor random networks, but networks with different statistical characteristics, which are called complex networks. P2P network, with features such as small world [2] and scale free [3], is a typical complex network. Related research [4] on Gnutella, the largest P2P application, found that 70% of Gnutella users rarely share resources and nearly 50% of the resources hits contribute by only 1% Gnutella users. This distribution of resources and node degree shows obvious scale-free feature. It is of great significance to analyze the distribution of P2P connections to design efficient resource search strategy, and control the amount of messages.

Using the power law distribution of P2P network to design efficient resource search is a new research field. Current resource search method in P2P network is designed to focus on the nodes, and refer to nodes' visit history and information of the neighbor nodes. These design principles increase the information processing of large numbers of additional nodes, and it is also difficult to measure the search efficiency and query costs. Our paper proposes an efficient search strategy, reliable random walk, on the scale-free feature of P2P network. And simulation experiments are designed to verify the analysis of the scale-free feature of P2P network and evaluate the performance of reliable random walk resource search.

Our Paper is organized as follow. Section 2 reviews the current related work, and in section 3 we give the assessment methods of scale-free feature of P2P network in reliable random walk. A detailed description of the proposed resources search strategy, reliable random walk, is given in section 4. Section 5 is for simulation and the experimental results are analyzed. Finally, in Section 6 we give conclusions and prospects of future work.

II. Related works

Currently, unstructured P2P network is widely used in P2P applications. Gnutella [5], which is a pioneer of P2P applications, uses flooding mechanism to discover shared resources in network. Flooding method is simple and easy to follow, however it results in too much search messages. Therefore, there have been many improved algorithms proposed to reduce the search costs caused by flooding method. Reference [6] was the first introduction of improved search algorithms for unstructured P2P network, such as Iterative Deepening, Directed BFS and Local Indices. Iterative Deepening determines whether to continue forwarding based on the results of each query,

doi:10.4304/jnw.5.11.1365-1372

which will increase the query delay due to the need to wait for each query results. In order to reduce the amount of query messages, Directed BFS selects a subset of neighbor nodes to forward search messages rather than forward query messages to all neighbors. However, it will increase the node processing cost and result in the aggregation of query requests to the excellent service nodes. In Local Indices, nodes store the information of resources in a radius neighbors. So, it only need make responses in partial TTL rounds to reduce search costs. But it requires nodes to store a large amount of resource information. Other resource search strategies of unstructured P2P network also include Adaptive Probabilistic Search (APS) [7], Random walk [8], PeerRank [9], Freenet [10], assist P2P search [11], Scalable Query Routing [12], etc.

A variety of resource search strategies in unstructured P2P network are basically designed based on nodes' history behavior, which needs additional statistical information. It makes against the stability of the routing efficiency and is difficult to fit the dynamic changes of P2P network. If resource search strategy is designed based on the overall dynamic characteristics of P2P network, it would be helpful to improve the search efficiency and overall performance of P2P network.

Current related study on P2P network shows that P2P network has an obvious scale-free property. M. Ripeanu and others used Crawler to make statistical analysis on the Gnutella network and found that node-degree distribution obviously comply with power law distribution. When E. Adar and other people research the Gnutella network Free Riding [4] phenomenon, they also have found that 70% of Gnutella users do not share any resources. While 25% of the nodes in the network handled 98% of the resource search requests. It also fully demonstrates that resource distribution and the search processing have a clear scale-free feature follow power law distribution, which has been found in many other study of P2P network applications [6] [7] [8]. Taking advantage of the scale-free feature of the P2P network as optimizing tools to improve the resource searching efficiency in P2P network is becoming a new research direction. However, at present the field of making effective use of scale-free feature and other characteristics of complex network to achieve measurable, efficient and stable network resource search is still a blank, which is the main motivation of our paper.

Nima Sarshar and others designed percolation search method [14] in complex network environment based on its scale-free feature and the percolation theory, which was also extended [15] into P2P network applications. This percolation search query is efficient and stable with the advantage that nodes need not to store large amounts of additional information. However, the percolation search needs resource redundancy for communication and construction of the suitable distribution structure. At present there are some resource search strategy [12] [16] use scale-free feature of P2P network to optimize resource search, but they are just qualitative using scale-free character, whose search efficiency is not stable with the dynamic changes of the peer and resource distribution. So it is also difficult to measure their efficiency.

We analyze the scale free distribution of P2P network, and design reliable random walk resource search strategy in unstructured P2P network environment based on the network inherent distribution character. Compared with similar improved random walk search algorithms [16] [17] and other unstructured P2P network optimizing algorithms, in reliable random walk no additional storage of statistical information is required for routing forwarding, and search efficiency is reliable and measurable. What is more, it also has high search efficiency, low network cost and easy to be extended on other network applications. Moreover, the reliable random walk algorithm also has a good reference value for other complex network researches, such as ad hoc, sensor networks, grid, etc.

III. Scale-free property assessment of P2P network

P2P network is a kind of complex network which is found with apparent scale-free properties in network connection by related research [13] [18]. That is, the network connection is re-tailed, and the majority of the nodes have only a few connected edges, but there are little nodes having a large number of connected edges. The scale-free property of complex networks is considered that the node degree is subject to the so-called power-law distribution. Namely, the proportion of nodes with the degree K in networks is showed in Formula (1), where A is the ratio constant and τ is the power-law distribution index.

$$P(k)=Ak^{-\tau} \tag{1}$$

The main content of our paper is to design an efficient, stable and measurable resource search strategy utilizing scale-free property in P2P network. Accurate assessment of the network's scale-free distribution is very important for our research. As P2P network is a highly dynamic one and nodes act of great autonomy. We use sample analytical tools to evaluate the P2P network power law distribution. We set up a scale-free assessment module over P2P network, which is consist of network connection acquisition sub-module and scale-free distribution analysis sub-module, which is responsible for statistics on the nodes' connection distribution in P2P networks. The system architecture is showed in Figure 1.

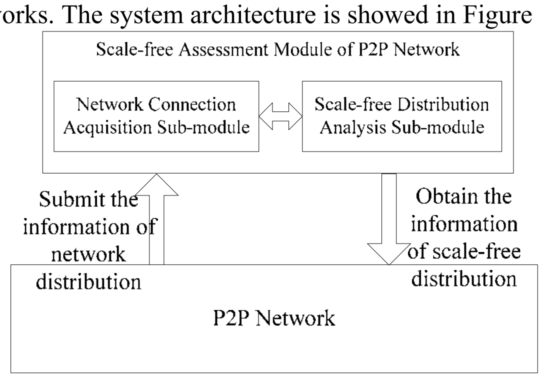


Figure 1. Scale-free analysis for P2P network

Be different from the centralized network resource search similar to Napster, this scale-free analysis module in reliable random walk is responsible only for the connection status of nodes in networks and scale-free analysis as well, which is not involved in the resource search. Network node periodic visits the scale-free analysis module to update the local network connection situation, getting the analysis of overall scale-free distribution. The study of Gnutella [13] found that the network connection is in strict conformity with the power-law distribution of Formula 1 in initial P2P network applications. However, with the extension of network application protocols, nodes which have small network degree (less than a certain threshold K) no longer meet this power-law distribution, while the nodes with larger network connectivity (greater than the threshold K) still meet it well. So, the presentation of the power-law distribution in Formula 2 better reflects the real scale-free distribution in P2P network. And it is also the basis of scale-free analysis in reliable random walk. Since the search requests obtain responses mainly in those highly connected nodes through the P2P resource searching process. The distribution of the highly connected nodes can still be described and analyzed through the power-law distribution are the main concern in reliable random walk.

$$P(k)=Ak^{-\tau},\ \tau>0,\ k\geq K \qquad (2)$$

The scale-free network analysis module may calculate the proportion $P(k)$ that nodes of different degree get through the network connection information reported by nodes. And then uses the function approximation method to determine the scale-free parameter A and τ as well as the scale-free phenomenon inflection point K based on Formula 3. Assessment methods are described below. First, assign u=ln$y$, v=ln$x$, and then transform Formula 3 to Formula 4 as a straight line form. Afterwards, take the advantage of the statistical data in network connection obtained to determine K, the inflection point of scale-free phenomenon. And next, use the least-squares fitting a straight line to determine the power-law distribution parameters A and τ in range of the scale-free distribution interval.

$$Ax^{-\tau},\ \tau>0,\ x\geq K \qquad (3)$$

$$u=\ln A-\tau v \qquad (4)$$

In the scale-free evaluation of reliable random work, a so-called lazy synchronization mechanism is in use to collect the distribution of nodes connectivity. As the scale-free statistical analysis module does not take the initiative to obtain the node-degree information but by the network nodes periodic report the local network connection status. While a new node joins the P2P network, it would have access to the statistical analysis module for the current network scale-free distribution and establish a stable network connection, then notify the module its connectivity degree. Using this synchronization mode can greatly reduce the amount of network message caused by simultaneous network connection status. Moreover, statistical analysis modules can be relatively real-time obtain network connection status.

This method ensures the accuracy of the scale-free analysis, thus the follow-up resource search strategy design is reasonable. Finally, the simulation experiments will be designed to evaluate the correctness of scale-free assessment strategies in reliable random work.

IV. RESOURCE SEARCH IN P2P NETWORK BASED ON SCALE-FREE PROPERTY

Reliable random walk search strategy which is based on network scale-free distribution, allocates resource hitting probability to each random walk search message to achieve the scalability and reliability of resource search. The resource search messages of reliable random walk are generated and distributed entirely based on the P2P network power law distribution, which can effectively improve the resource search efficiency and minimize network cost for unstructured P2P network.

In reliable random walk, nodes visit scale-free analysis module periodically to obtain scale-free distribution of the network, including the scale-free distribution parameters A, τ and the scale-free phenomenon inflection point K. Resource search strategy is designed based on the scale-free distribution of the network. Figure 2 shows the structure of reliable random walk's resource search query message. When resource search request is initiated by the node, the query condition *Query* is generated firstly. Related research [4],[13] on P2P network application found that resource search request in P2P network is responded mostly in the nodes with high node-degree. Therefore, in reliable random walk resource search, $Degree_{high}$ is set according to the network distribution. The central nodes are defined as nodes whose connectivity is not less than $Degree_{high}$. When a search request message reaches a central node, it can be satisfied since the central nodes contain most of the resources information in the local network. The search request should not continue forward, which minimize the amount of network messages while the resource search efficiency does not be reduced. In a search request message, q represents the probability of hitting the central nodes undertaken by this message, which is used to quantify the search efficiency in reliable random walk. This probability is assigned to each resource search message to ensure the searching efficiency. The assignment of hitting probability and the distribution of search message will be further analyzed in next section. TTL represents the maximum radius of search request spreading.

Query	$Degree_{high}$	q	TTL	......

Figure 2. Resources search query message of reliable random walk

Suppose the parameters of node P in P2P network obtains scale-free distribution is A and τ, scale-free inflection point is K. In reliable random walk, the central node, namely $Degree_{high}$, is determined by an adaptive strategy. $Degree_{high}=Degree_{MAX}/2$ in initialization and it can be updated by nodes in range of [K, $Degree_{MAX}$] according to the query situation. If a node wants to get information more, it can assign a bigger $Degree_{high}$, and vice-versa. When a query source node P generates a query condition *Query* and expects query message to hit the central node with probability Q_{ALL}, it needs to calculate Q_{high}, the proportion of central nodes according to the network distribution. Formula 5 shows the proportion of central nodes in P2P network, making use of scale-free distribution property of P2P network showed in Formula 2.

$$Q_{high} = \sum_{k=Degree_{high}}^{Degree_{\max}} Ak^{-\tau} \tag{5}$$

How can ensure to hit the central nodes with expected probability Q_{ALL}, it is the core issue of reliable random walk to forward resource search request depends on the power law distribution of P2P network. Suppose that the largest transmission radius of query message is *TTL*, and the probability of search request hit the central nodes in each *TTL* rounds. Formula 6 is necessary in order to meet the overall resource search hit rate. Suppose that the hit probability of each *TTL* round resources search is roughly equal, then it can be considered that the central node hit rate of each *TTL* round Q meets Formula 7. Source node of the search request P calculate how many search request is needed based on the central node hit rate of each *TTL* round Q. Suppose the network connections of P is k, then in order to meet the hit rate Q, the amount of search request r is shown as Formula 8 and central nodes hit rate that every search request need to take on q is shown as Formula 9. Query request messages carry their own hit rate q to forward, which is keeping fulfilling in the following forwarding *TTL* to achieve the reliability and metrizability of overall resource search in reliable random walk.

$$1-(1-Q_1)(1-Q_2)\ldots(1-Q_{TTL}) \geq Q_{ALL} \tag{6}$$

$$Q \geq 1-\sqrt[TTL]{1-Q_{ALL}} \tag{7}$$

$$1-(1-Q_{high})^r \geq Q \Rightarrow r \geq \frac{\ln(1-Q)}{\ln(1-Q_{high})} \tag{8}$$

$$q = 1-\sqrt[r]{1-Q} \tag{9}$$

In reliable random walk, the total query hit is break down into each random walk search request. Unfortunately, a sufficient number of random walk ($k<r$) is not allowed according to the connectivity degree k of some nodes. It would undermine the stability and reliability of the overall resource search efficiency of reliable random walk. The following context of the specific query process is to analyze how to ensure the system query efficiency in this case.

Resource search process is shown in Figure 3, in which a search request source node P expects that the search request hits the central node in a probability of 90%. Firstly, calculate the proportion of the central node by Formula 5 Q_{high}=5.26%($Degree_{MAX}$=10), and set the maximum search radius TTL = 6, then in order to achieve the total search hitting rate as 90%, for each *TTL* round the probability of hitting the central node needs to be over 31.87%(Q>31.87%). When a node P delivers a search request, in order to satisfy the first-round search hitting probability the numbers of random walk r should be not less than 7.10 as computed by Formula 8 (r>7.10). Unfortunately the network connection degree of the node P can not meet the requirement of publishing 8 random walks. In reliable random walk, when the search requests published by a node can not meet the requirements of the central node hitting probability, the node forwards or transits the query messages by its own network connection number k, during which each random walk carries the hitting probability of meeting the expectations but not make the hitting probability in each round as the allocation criterion. The search hitting probability will be submitted completed in the follow-up *TTL* round to ensure the overall source searching efficiency.

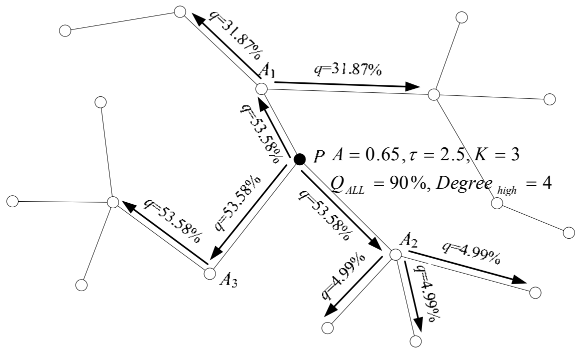


Figure 3. Reliable search based on power law distribution

As shown in Figure 3, node P releases 3 random walks, using Formula 10 to calculate the hitting central node rate of each message taking q = 53.58%, putting q in each of the random walk. Based on the local network scale-free distributed property, nodes which received query messages use a similar strategy to calculate the number of random walk to be transmitted. In order to complete the hitting rate of each search message. Supposing the neighbor nodes obtain the same scale-free distribution as P, then in the remaining *TTL*-1 round, the hitting rate of each message transmitting round is expected to reach $1-(1-Query_{high})^r \geq Q \rightarrow r \geq \frac{\ln(1-Q)}{\ln(1-Q_{high})}$, and the number of random walk needs to be transmitted is $q = 1-\sqrt[r]{1-Q}$.In Figure 3 node A_2 can transmits 3 query requests to meet the hitting rate of last round of random walk query. And each random walk possesses $q' = 1-\sqrt[r]{1-Q_{ALL}}$ hitting rate of the central nodes. The nodes A_1 and A_3 in Figure 3 can not transmit three random walks, therefore nodes transmits the random walk with their best capacity, and calculate the query hit rate of every random walk should take by Formula 11. The following query process is similar.

Our design is also reasonable, the amount of transmitted messages is calculated by Formula 8, while in Formula 8, the central node hitting rate Q of each round is supposed are the same in every *TTL* rounds. However, when the *TTL* is small, the quantity of query message volume is small, and the total node degree is low, so it's impossible to obtain the same hitting rate of central nodes as the following rounds. For this reason, in the beginning of reliable random walk search process, search request information always determines central node hitting rate of every query round according to the expected hit rate Q_{ALL}.

So, nodes try their best to forward search request information. In the following *TTL* searching process, the amount of query messages is determined to minimize network messages according to hitting rates of each round. Therefore, reliable random walk is still efficient to the whole search as long as it can achieve the hitting rate of every search message. In the simulation experiment we will analyze the situations which evaluate the resource search message hitting the central nodes in the reliable random walk.

$$q' = 1 - \sqrt[r]{1 - Q_{ALL}} \tag{10}$$

$$q' = 1 - \sqrt[r]{1 - q} \tag{11}$$

In reliable random walk resource searching, when *TTL*=0 or the central node is hitting, the query message stops forwarding. The former stop condition is for the reason of controlling the query radius and the later one is in the consideration of reducing the amount of query messages under the precondition of ensuring the query efficiency. If the probability of a central node appears is Q_{high}=5.26%, then the probability that a query message in the query range of *TTL*=6 hits two or more central nodes is

$$1 - (1 - Q_{high})^{TTL} - C_{TTL}^{1}(1 - Q_{high})^{TTL-1} Q_{high} \approx 3.60\%$$

. Such a probability is very low, which is the main reason why the query message stops forwarding when it has hit a central node.

The reliable random walk resource searching makes the searching hit and random walk forwarding by using the scale-free characteristics of P2P network. When the connectivity degree of a node cannot ensure the query messages that have been sent reach the excepted hitting probability, the messages will be forward in the utmost. Otherwise, the messages will be forward in a number that can ensure the expected search hitting probability.

From the analysis of the retrieval process, it can be found that during the spreading of the query message, the query hitting probability of every random walk drops rapidly. In the following forwarding, no more query messages are needed. Such a resource searching strategy is conducive to rapid discovery of the central node and then response to the searching request. Besides, it can also control the amount of the messages in the network. Such a method is conducive to the stability of the unstructured P2P network resource search. In the following experiment section, related experiments are designed to evaluate the performance of reliable random walk resources searching.

V. Simulation Experiments and Results Analysis

The following simulation experiments are designed to simulate reliable random walk resource search strategy and evaluate its performance. Statistic of its scale-free analysis and efficiency in resource search has been made, and we compare reliable random walk with similar resources search algorithm to analyze the simulation results.

A. Experiment Setting

Reliable random walk is a measurable and efficient resource search algorithm base on unstructured P2P network's scale-free distributed properties. In the simulation experiments, using PLOD algorithm [19] we get a P2P network scale-free topology with τ=2.5-3.0, A=0.60-0.75 to simulate the actual P2P network application environment, and design simulation network scale-free turning point K = 3, that is nodes with $k \geq 3$ meet scale-free distribution. The network size of simulation experiments is 1000-5000, and the maximum node connectivity degree $Degree_{MAX}$=10-20. Experiment 4.2 and 4.3 test and evaluate the scale-free analysis accuracy and shooting proportion of central nodes in the reliable random walk, which is implemented by JAVA. Experiment 4.4 tests the resource search strategy of reliable random walk, besides, compares the reliable random walk with the Gnutella and k-random walk search strategy, which are built by PeerSim [20]. Tests are in the same experimental environment and network parameters.

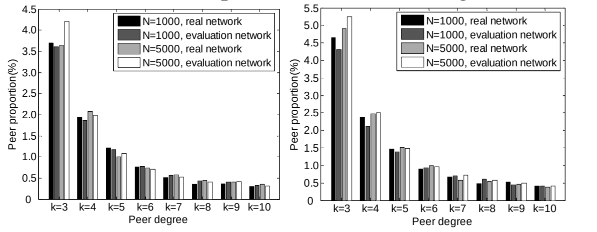
All simulation experiments are performed on a single PC, with configuration of CPU P4 2.8GHz, Memory 1GB, and Windows XP operating system. Table 1 shows the basic parameter settings of simulation experiments.

TABLE I. Parameter and settings in the simulations

	Parameter meaning	**Value**
N	Network size	1000, 5000
A	Proportional coefficient of scale-free distribution	0.60, 0.75
τ	Index of scale-free distribution	2.5, 3.0
K	Inflection point of scale-free phase change	3
TTL	Maximum forwarding hops of query messages	6
$Degree_{MAX}$	Maximum connectivity degree of nodes	10, 20
Q_{ALL}	Expected hitting proportion of central nodes	80%, 90%

B. Scale-free analysis of P2P network

Reliable random walk resources search expands entirely based on the scale-free distribution properties of P2P networks. The accuracy of the network scale-free analysis is very important to the resource search, so we firstly design the simulation test to detect the accuracy of scale-free analysis in the reliable random walk. Maintain the network size as N, and all nodes join in and leave the network randomly every 300 seconds, and the scale-free analysis module reports local network connection information every 600 seconds. In practical P2P networks, it is difficult to collect network connection information of all nodes, so in the simulation experiments, the scale-free analysis module just selects only 20% network connection information of the latest nodes for network scale-free property analysis. Simulation experiment tested in different scale-free distribution and network scope to compare the difference on the network practical node degree distribution and node degree distribution in the P2P network scale-free module with scale-free analysis. Experiments last 5 hours, we collect the data of current node distribution in the network and evaluation data of scale-free module analysis every 5 minutes, then average and compare both of the results. The comparison is shown in Figure 4.



(a)τ=2.5, A=0.60 (b)τ=2.5, A=0.75

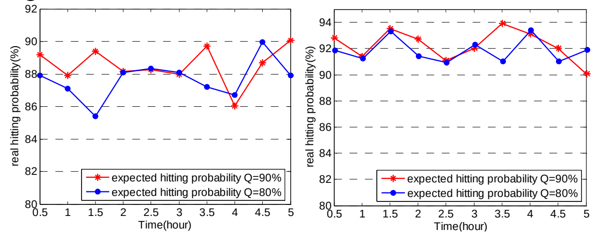
(c)τ=3.0, A=0.60 (d)τ=3.0, A=0.75

Figure 4. Analyses of Peer-to-Peer network power law distribution

In the reliable random walk, resource search shooting focuses on the central nodes with a high degree of network connectivity, therefore, scale-free analysis module makes use of the collected node data of network connection to analyze scale-free phase transition point K, as well as scale-free parameters A and τ, to determine the network scale-free distribution, and predict the distribution of nodes with a high network connection level (k>K). The experiment results show that by evaluating connection distribution of P2P network with scale-free analysis module and computing the distributed proportion of high connectivity nodes in the network according to the evaluation result, we can make the predicted distribution almost the same as the actual node distribution in the network. It can be found in the experimental results that the design of P2P network scale-free analysis strategy is effective and nodes can correctly understand other nodes distribution in the network with the help of analysis results, so as to ensure the correctness and effectiveness of following resource search strategy based on scale-free properties.

C. Query hitting probability of central nodes

Reliable random walk resource search is designed based on the reliable hit central node, so it is of great importance for the stability and scalability of reliable random walk's efficiency in searching resources. We care about whether reliable random walk resource search request message can hit the central node of the P2P network stably and reliably according to expected probability. We design simulation experiments to study the issue under different network conditions. The largest connectivity degree of the network nodes $Degree_{MAX}$=10, and the simulation experiment determine the central node in accordance with $Degree_{high}$= $Degree_{MAX}$ /2=5. The network size is maintained at N=5000. Every 60 seconds network nodes make a search request to the network with the expected probability Q_{ALL}, TTL = 6, and obtain scale-free distribution conditions every 600 seconds by accessing the scale-free analysis module. After 5 hours, we make a statistics of network query request's hitting probability of the central node according to the network run-time. The simulation experiment results are shown in Figure 5.



(a)τ=2.5, A=0.60 (b)τ=2.5, A=0.75

(c)τ=3.0, A=0.60 (d)τ=3.0, A=0.75

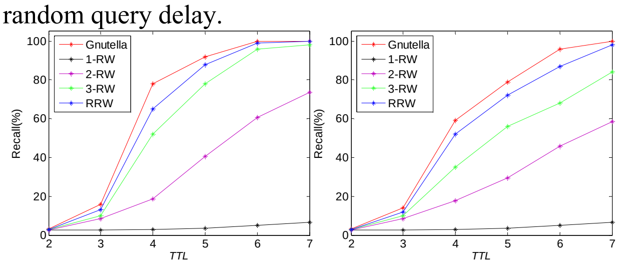
Figure 5. Rate of hitting high degree nodes

Experimental data shows that reliable random walk resources search strategy can ensure that the search requests can hit the network central nodes within the expected query hit probability Q_{ALL} in the query scope TTL limited. The hit probability of central node is slightly lower than expectations when the proportion of the network central node is very low(in Figure 5c situation, the ratio of the central node is 1.19%). Analyzing the results, when we query the nodes whose continuous hit rate in querying messages is extraordinary low in this case, the query message cannot be spread effectively, leading to a result that the central node cannot be hit effectively. In this case reliable random walk is almost equivalent with flooding messages. Nodes which have received query messages will transmit them with maximum capability, and their failure to hit the central node is result from network topology. Therefore, reliable random walk is a kind of adaptive resource search strategy, the nodes can set query request according to their own needs and the network scale-free distribution so as to achieve an optimal resource search. Next we will compare reliable random walk with other similar P2P network resources search strategies and make a further analysis about the search performance of reliable random walk.

D. Search efficiency of Reliable random walk

Reliable random walk is a search strategy for unstructured P2P network resources based on networks with scale-free characteristic, of which query efficiency can be measured and assured. In order to test the resources search efficiency of reliable random walk, we design some simulation experiments to make a comparison with other resources search strategy such as Gnutella, k-random walk. We use PeerSim [20] to build Gnutella, k-random walk, and reliable random walk network protocol, and we validate the efficiency of their routing in the same experimental environment. The network topology of the simulation experiments is a Power Law distribution with different distribution conditions. Its scale-free phase-change point K=4, namely, the nodes whose k≥4 meet the scale-free distribution. The network size N=5000 and $Degree_{MAX}$=20. The resources each node stored accord with the Zipf distribution that α=1.2, $Document_{MAX}$=200, and the number of search request each nodes initiated accord with the Zipf distribution that α=1.0, $Query_{MAX}$=100.

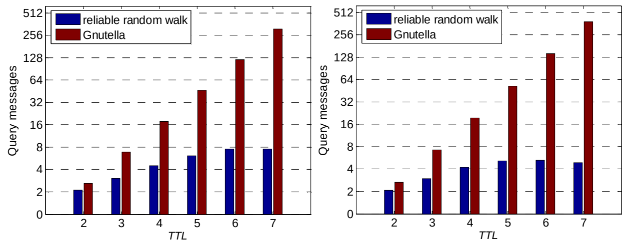
Firstly, simulation experiments are designed to compare the recall of three kinds of resources search strategy - Gnutella (flooding), k-random walk (k-RW for short, k = 1, 2, 3), and reliable random walk (RRW for short) under the condition of different P2P networks scale-free distribution. Reliable random walk node search network resources and make statistics of network average recall rate in accordance with TTL = 7 and expected probability of 90%. The results of experiment are shown in Figure 6. The experimental results show that recall rate of reliable random walk can be basically guaranteed in about 80% when TTL> 5, and it can be nearly 90% when TTL> 6. Reliable random walk is designed based on the distribution of P2P network resources, the experiment data show that reliable random walk can achieve a equivalent recall rate with the flooding strategy of Gnutella as the network changes in the distribution of scale-free, and it shows better adaptability for the changes of peer networks scale-free distribution compared with k-random walk. Meanwhile reliable random walk gets a higher recall rate compared with k-random walk when the TTL is small, which is conducive to improving the k-random query delay.



(a)τ=3.0, A=0.75 (b) τ=2.5, A=0.60

Figure 6. Recall of Gnutella, k-random walk and reliable random walk

Because the amount of searching messages is also a quality index to focus on in resource search, the simulation experiments was designed to compare the amount of query messages among different networks with the condition of scale-free distribution. As the messages of k-random walk is constant with the change of TTL, only the comparison of the query message amount between Gnutella and reliable random walk is necessary. In the simulation experiment reliable random walk node searches network resources in accordance with TTL = 7, and the expected probability is 90%, then we make statistics of query messages amount of each TTL message forwarded round. The experimental results are showed in Figure 7. From the experimental data it can be found significantly that as the TTL query messages increasing the amount of query message of reliable random walk is far less than the amount of Gnutella's query message, while the recall of reliable random walk was almost equal with Gnutella. It is also found in the simulation experiment that as the TTL increases the query message amount of reliable random walk is not increasing sharply. A in-depth analysis revealed that the query message is not forwarded after the central node is hit in query process, which greatly reduces the growing query message later, and because the central node stores a large number of network resource information, queries can be correctly responded, which ensures that the query efficiency will not be affected.



(a)τ=3.0, A=0.75 (b) τ=2.5, A=0.60

Figure 7. Query messages of Gnutella and reliable random walk

VI. SUMMARY AND FUTURE WORK

This paper presents reliable random walk, a resource search algorithm for unstructured P2P network. It is based on the scale free character which is an inherent character of P2P network. The main advanteges of reliable random walk is that its searching efficiency is reliable and metrizable. Reliable random walk predicts the proportion of the central nodes, uses query messages hitting the central node to quantify the resource search efficiency of the P2P network. Each query message in the network brings the quantitative hitting probability, which realizes the metrizability and high reliability of P2P network resource search.

Compared to other similar P2P network resources search algorithms, reliable random walk has lower maintenance costs, need not to store any additional supporting information, and does not require statistical analysis on the information of the history search requests. In addition, reliable random walk is easy followed in unstructured P2P network and easy to implement. During the experiment we found that the deterioration of local network condition, such as low local network connectivity, will cause some reliable random walk query failure. Therefore, in future research work, we'll consider the introduction of local scale-free feature analysis of the node to enable node's control of local P2P network's distribution. In this way query messages can be adjusted adaptively according to the distribution of local network, which will make the network query scalable.

ACKNOWLEDGMENT

This research is partially supported by Natural Science Foundation of China under Grant No. 70901060, China Postdoctoral Science Foundation under Grant No. 20100471145, the Fundamental Research Funds for the Central Universities under Grand No. 6082024, National the Natural Science Foundation of Hubei Province under Grant No. 2009CDB304.

REFERENCES

[1] P.Erdos, A.Renyi. On the Evolution of Random Graphs. In: Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 1960, 5:17-61

[2] D. Watts, S. H. Strogatz. Collective Dynamics of Small World Networks. Nature, 1998, 393: 440-442

[3] Reka Albert, Albert-laszlo Barabasi. Statistical Mechanics of Complex Networks. Reviews of Modern Physics, 2002, 74: 47-97

[4] Eytan Adar, Bernardo A. Huberman. Free Riding on Gnutella. Xerox PARC, 2000

[5] Gnutella. http://www.gnutella.com

[6] Beverly Yang, Hector Garcia-Molina. Improving Search in Peer-to-Peer Networks. In: Proc. of the 22nd International Conference on Distributed Computing Systems (ICDCS), 2002, pp. 5-14.

[7] Dimitrios Tsoumakos, Nick Roussopoulos. Adaptive Probabilistic Search for Peer-to-Peer Networks. In: Proc. of the 3rd International Conference on Peer-to-Peer Computing (P2P), IEEE, 2003, pp. 102-109.

[8] Qin Lv, Pei Cao, Edith Cohen. Search and Replication in Unstructured Peer-to-Peer Networks. In: Proc. of the 16th ACM International Conference on Supercomputing (ICS), New York, ACM Press, 2002, pp. 84-95.

[9] Feng Guo-Fu, Mao Ying-Chi, Lu Sang-Lu, Chen Dao-Xu. PeerRank: A Strategy for Resource Discovery in Unstructured P2P Systems. Journal of Software, 2006, 17(5): 1098-1106.

[10]Ian Clark, Oskar Sandberg, Brandon Wiley, Theodore W. Hong. Freenet: a Distributed Anonymous Information Storage and Retrieval System. In: Proc. of the Workshop on Design Issues in Anonymous and Unobservability. Heidelberg: Springer-Verlag, 2000, pp. 311-320.

[11]Rongmei Zhang, Y. Charlie Hu. Assisted Peer-to-Peer Search with Partial Indexing. In: 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), New York: IEEE Press, 2005, 1514-1525.

[12]Abhishek Kumar, Jun Xu, Ellen W. Zegura. Efficient and Scalable Query Routing for Unstructured Peer-to-Peer Networks. In: 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), New York: IEEE Press, 2005, 1162-1173.

[13]Matei Ripeanu, Ian Foster, Adriana Iamnitchi. Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design. IEEE Internet Computing Journal, 2002, 6(1): 50-57

[14]Nima Sarshar, Oscar Boykin, Vwani Roychowdhury. Scalable Percolation Search on Complex Networks. Theoretical Computer Science. 2006, 355(1): 48-64.

[15]Nima Sarshar, P. Oscar Boykin, Vwani P. Roychowdhury. Percolation Search in Power Law Networks: Making Unstructured Peer-to-Peer Networks Scalable. In: Proc. of the 4th International Conference on Peer-to-Peer Computing(P2P), IEEE, 2004, pp. 2-9.

[16]Nabhendra Bisnik, Alhussein A. Abouzeid. Optimizing Random Walk Search Algorithms in P2P Networks. The International Journal of Computer and Telecommunications Networking. 2007 51(6): 1499-1544.

[17]Ming Zhong, Kai Shen. Popularity-Biased Random Walks for Peer-to-Peer Search under the Square-Root Principle. In: Proc. of the 5th International Workshop on Peer-to-Peer Systems (IPTPS), Santa Barbara, CA, 2006.

[18]Stefan Saroiu, P. Krishna Gummadi, Steven D. Gribble. A Measurement Study of Peer-to-Peer File Sharing Systems. In: Proc. of the Multimedia Computing and Networking (MMCN), SPIE/ACM, San Jose, CA, USA, 2002, pp. 18-25

[19]Palmer CR, Steffan JG. Generating Network Topologies that Obey Power Law. In: Proc. of the Global Internet Symposium (Globecom), San Francisco, IEEE, 2000, pp. 434-438.

[20]PeerSim. http://peersim.sourceforge.net/

Wei Song received the BS degree in Mechanical Science and Engineering from Huazhong University of Science and Technology, China in 2001, and Ph.D. degree in Computer Science from Huazhong University in 2008. He is currently a lecture and post-doctor in the School of Computer at Wuhan University. His main research interests include peer-to-peer network, distributed system, and distributed system security.

Wenbin Hu Associate professor worked in Wuhan University since 2006. He got his Ph. D degree in 2004. His main research interests include intelligent simulation and optimization, multi-agent system and swarm intelligent algorithm.

Zhengbing Hu received B.E., M.E. and Ph.D. degree in National Technical University of Ukraine. He is currently a lecture in Huazhong Normal University. His current research interests include network security, intrusion detection system, artificial immune system, data mining etc.

Xi Zeng received his bachelor's degree in Computer School from Wuhan University in 2010. Now he is pursuing his M.S. degree in Department of Computer Science and Engineering at the Chinese University of Hong Kong, Hong Kong, China. His main research interests include distributed systems, peer-to-peer network.