2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing

Resource Search in Peer-to-Peer Network Based on Power Law Distribution

Wei Song[1*], Xi Zeng[1], Wenbin Hu[1], Yiting Chen[2], Chuanjian Wang[3], Fangquan Cheng[1]
[1]Computer School, Wuhan University, Wuhan 430072, China
[2]JiangCheng College, China University of Geoscience, Wuhan 430200, China
[3]College of Information Science and Technology, Shihezi University, Shihezi 832003, China

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

Keywords: Peer-to-Peer; reliable random walk; complex network; scale free; power law distribution

I. INTRODUCTION

A large number of complex systems in nature, such as the Internet, transport system, social networks and so on, can be described by network. The earliest research on network is based on Graph Theory. In the mid-twentieth century, Erdos proposed that the network connection is random. Based on this standpoint, the random network model [1] is established, which had been the theoretical basis of real network research for a long time. Recently, 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 has obvious features of small world [2] and scale free [3], so it 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 are contributed by only 1% Gnutella users. This distribution of resources and node degree has obvious scale-free feature. Therefore, it is of great significance to design efficient resource search strategy by analyzing the distribution of P2P links.

The available resource search in P2P network is designed to focus on the nodes, and refer to the information of the neighbor nodes. These design principles have to greatly increase the processing of nodes, and it is difficult to measure the search efficiency and costs. Comparing with the existing researches, our paper proposed an efficient search strategy, reliable random walk, based on the scale-free distribution of P2P network which is an inherent character. Moreover, we design experiments to evaluate its performance.

Our Paper is organized as follow. Section 2 reviews some related work, and in section 3 we give the assessment methods of P2P network scale-free distribution. A detailed description of the reliable random walk method is given in section 4. Section 5 is for simulation and the experimental result analyze. Finally, in Section 6 we give conclusions and future work.

II. RELATED WORKS

Currently, unstructured P2P network is widely used in P2P applications. Gnutella, 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 costs. Therefore there have been many improved algorithms to reduce the search costs caused by flooding method. Reference [5] was the first introduction of improved search algorithms including Iterative Deepening, Directed BFS and Local Indices. Other improving search strategies of unstructured P2P network also include Adaptive Probabilistic Search (APS) [6], Random walk [7], PeerRank [8], assist P2P search [9], Scalable Query Routing [10], etc. All of them are based on nodes' behavior, which needs additional statistical information. It is difficult to adapt to the dynamic changes of network and huge network size.

M. Ripeanu uses crawler to make statistical analysis on P2P network found that node-degree distribution has obvious scale-free feature[11].While E. Adar researches the Gnutella network Free Riding phenomenon [4], he also found that 70% Gnutella users do not share any resources, while 25% nodes handled 98% of the resource search requests. So, taking advantage of the P2P scale-free feature to improve resource search in P2P network is a new research idea.

Nima Sarshar proposed percolation search [12] in complex network based on its scale-free feature and percolation theory, which was also extended into P2P network [13]. This percolation search query is efficient and stable without storing large amounts of additional information. However, the percolation search needs additional resource copy and a suitable network structure. So, the percolation search is difficult to follow. Presently, there are some searching algorithms [10] [14] use scale-free feature to improve resource search efficiency. However, they just qualitative use scale-free distribution, their search efficiency is not stable with the dynamic changes of the network distribution.

* Corresponding author.
E-mail addresses: sw_cyt@126.com (Wei Song)

978-0-7695-4011-5/10 $26.00 © 2010 IEEE
DOI 10.1109/NSWCTC.2010.278

By analyzing the scale-free distribution of P2P network, our paper proposes reliable random walk, a new resource search algorithm. This algorithm has high search efficiency, low network cost and easy to be followed. Moreover, it has a great reference value for other resources search in complex networks, 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 [11]. It is found that the network connection is re-tail. And the majority of the nodes have only a few connections, only little nodes have a large number of connections. The scale-free feature of large-scale complex networks is considered that the node degree distribution follows the power-law distribution. The proportion of nodes with the degree K in networks is showed in Formula (1), where τ is the power-law parameter.

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

Our resource searching method is based on the network distribution, so it is important to accurately assess the network scale-free distribution. We use sampling analytical tools to assess the scale-free feature in P2P network. The system architecture is showed in Figure 1.

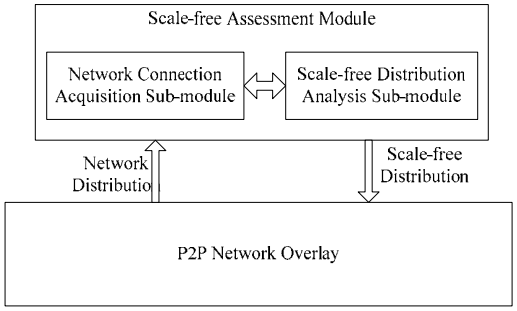


Figure 1. Scale-free analysis of P2P network

Be different from the centralized server, scale-free analysis module is responsible only for the connection status of nodes, which is not involved in resource search. Network node regularly visits the module to update the local connection situation and get the analysis results of overall scale-free distribution. The study of Gnutella [11] found that the network connection strictly follows the power-law distribution as Formula 1 in initial P2P applications. However, with the extension of P2P protocols, the nodes with small network degree (less than a threshold K) no longer meet this power-law distribution, while the nodes with larger network connection still meet it well. So, Formula 2 adapted by us can reflect the real scale-free distribution in P2P network better.

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

Our assessment methods are described below. First, assign $u=\ln P(k)$, $v=\ln k$, and then describe Formula 2 into Formula 3 as a line function. Afterwards, we use linear fit method to determine the power-law parameters A, τ, and threshold K.

$$u=\ln A-\tau v \tag{3}$$

In reliable random work, a lazy synchronization mechanism is used to collect the distribution of node connection. Nodes in P2P network just report the connection status when its connection changes. This synchronization mode can greatly reduce the messages caused by collecting network connection status.

IV. RESOURCE SEARCH BASED ON SCALE-FREE PROPERTY

Reliable random walk algorithm is based on network scale-free distribution to provide reliable search efficiency. It allocated search hit probability to each random walk message to achieve the scalability and reliability of resources search.

In reliable random walk, nodes visit scale-free analysis module periodically to get scale-free distribution information of P2P network. Figure 2 shows the structure of a reliable random walk's query message. Related researches [4] [11] on P2P network application found that search request in P2P network is responded mostly in the nodes with high node-degree. Therefore, we define the hot nodes as nodes whose connectivity are not less than $Degree_{high}$ which is based on network connection distribution. When a search request reaches a hot node, it can be satisfied since the hot nodes store most of the resource information local. So, the search request does not continue forwarding. This mechanism minimizes the query message amount while the resource search efficiency is ensured. In a search message, q represents the probability of hitting the hot nodes undertaken by this message. This probability is assigned to each query message to ensure the efficiency of total search.

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

Figure 2. Resources search query message of reliable random walk

Suppose the scale-free parameters that node P obtains is A and τ, scale-free inflection point is K. In reliable random walk, $Degree_{high}$, is set as $Degree_{MAX}/2$ in initialization and it can be adjusted by nodes in range of $[K, Degree_{MAX}]$ according to the query situation. The source node of a search request needs to analyze Q_{high} which dominates the probability of expection to catch hot nodes according to the network distribution. Formula 5 shows the proportion of hot nodes in P2P network based on scale-free property.

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

It is a core issue that how to ensure the hot node hitting probability of Q_{ALL}. Suppose that the largest transmission hops is *TTL*. Moreover, the probability of hitting the hot nodes in each *TTL* rounds is Q_i. To meet the overall hit rate Q_{ALL}, Formula 6 is necessary. Suppose that the hit probability of each *TTL* round is roughly equal, then it can be considered that the hot node hitting rate of each *TTL* round Q meets Formula 6'. Then source searching node P calculate how many search request is needed to satisfy Q. Suppose the network connections of P is k, then in order to meet the hitting rate Q the amount of search request r is shown as Formula 7. Moreover, each search request take on corresponding hitting

probability q is shown as Formula 8 to ensure the overall searching efficiency.

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

$$Q \geq 1-\sqrt[TTL]{1-Q_{ALL}} \quad (6')$$

$$1-(1-Q_{high})^r \geq Q \Rightarrow r \geq \frac{\ln(1-Q)}{\ln(1-Q_{high})} \quad (7)$$

$$q = 1-\sqrt[r]{1-Q} \quad (8)$$

In reliable random walk, the total hot node hitting is decompounded to each search request. However, some nodes' degree maybe not satisfy the sufficient number of random walk ($k<r$) messages. We analyze this situation as follow. P2P overlay is shown in Figure 3, in which a query source node P expects that the search request hits the hot node in a probability of 90%. Firstly, calculate the proportion of the hot 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 hot node needs to reach 31.87 percents(Q>31.87%). While node P delivers a search request, in order to satisfy the first-round searching probability, the numbers of random walk r should not less than 7.10. Unfortunately the network connection degree of P can not publish 8 messages. In this case, the node forward or transit messages as best as it can. Moreover, each message catches its hitting probability decompounded by overall hitting rate. And the search hitting probability will be satisfied in the continuous TTL round.

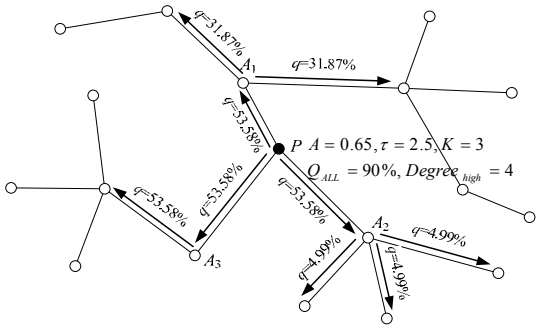


Figure 3. Reliable random walk search based on the scale-free property

As shown in Figure 3, P releases 3 random walks. And, each message takes q = 53.58% probability calculating by Formula 9. In the next transmitting, nodes received query messages use a similar strategy to calculate the number of forwarding messages. As the figure shown, in the remaining TTL-1 round, the hit rate of each search round should reach $1-(1-Q_{high})^r \geq Q \Rightarrow r \geq \frac{\ln(1-Q)}{\ln(1-Q_{high})}$, 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 messages to meet the hit rate of last round of random walk query, and each random walk takes q=1-(1-0.1423)$^{1/3}$=4.99% hitting rate. However, 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 hitting rate of every random walk by Formula 9'. The following query process is similar. Our design is reasonable. At the beginning transmitting round, the query message amount is small, and the total node degree is low, so it is difficult to satisfy the same hit rate as the following rounds. For this reason, in the beginning searching rounds, nodes try their best to forward search request information. In the following TTL search process, the amount of query messages is determined to minimize network messages according to hit rates of each round. Therefore, reliable random walk can support an efficient resource searching.

$$q' = 1-\sqrt[r]{1-Q_{ALL}} \quad (9)$$

$$q' = 1-\sqrt[r]{1-q} \quad . \quad (9')$$

In reliable random walk resources searching, when TTL=0 or the hot node is hit, the query message stops forwarding. The former stop condition is for the reason of controlling the range of the query and the later situation is in the consideration of reducing the amount of query messages as much as possible to ensure the query efficiency. If the probability of a hot 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 hot nodes is P=1-(1-Q_{high})^{TTL}-C^1_{TTL}(1-Q_{high})$^{TTL-1}$$Q_{high}$≈3.60%. Such a probability is low, which is the reason why the query message stops forwarding when hit the hot node.

From the analysis, it can be found that after the spread of the query message, the query hit 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 discover hot nodes. Besides, it can also control the amount of the messages in the network. Taking the advantage of the scale-free property of P2P network, reliable random walk implements a resource search method whose searching efficiency is measurable and scalable.

V. SIMULATION EXPERIMENTS AND RESULTS ANALYSIS

In this section, we design simulation experiments to evaluate the performance of reliable random walk algorithm.

A. Experiment Setting

In the experiments, we uses PLOD algorithm[16] to build scale-free network topology, and the scale-free change point K =3. The simulation network size is 1000-5000. Moreover, the maximum node degree $Degree_{MAX}$=10-20. The experiments are built over PeerSim[17]. Table 1 shows other experiment settings.

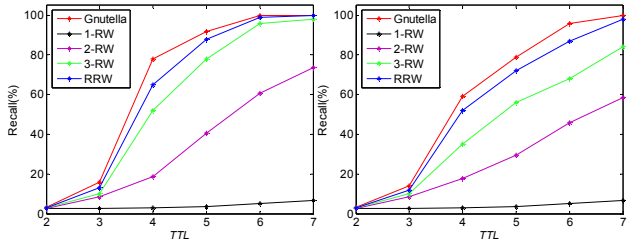
TABLE I. PARAMETER AND SETTINGS IN THE SIMULATIONS

	Parameter meaning	**Value**
TTL	Maximum forwarding hops	6
Q_{ALL}	Expected hitting proportion of hot nodes	80%, 90%

B. *Search efficiency of Reliable random walk*

We compare reliable random walk with other P2P search algorithm such as Gnutella and *k*-random walk in same environment.

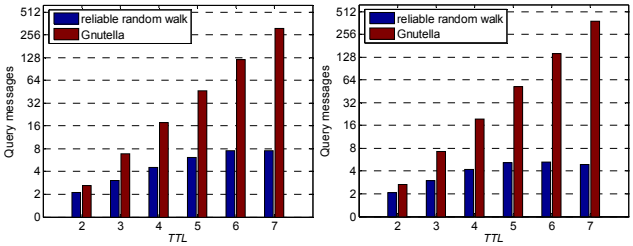
Firstly, experiments evaluate the recall rate of three searching algorithm: Gnutella, *k*-random walk, and reliable random walk. Reliable random walk node searches resources with expected probability of 90%. The experimental results in Figure 6 show that recall of reliable random walk can reach about 80% when *TTL*>5, and it can be nearly 90% when *TTL*=6. Moreover, the experimental data show that reliable random walk achieve an equivalent recall with the flooding strategy in Gnutella. Compared with *k*-random walk, reliable random walk shows better adaptability and gets a higher recall.



(a)τ=3.0,A=0.75,$Degree_{MAX}$=20 (b)τ=2.5,A=0.60,$Degree_{MAX}$=20

Figure 4. Recall of Gnutella, k-random walk and reliable random walk

Considering that the query message amount is also a performance of searching algorithm. We design simulation experiments to compare the message amount between reliable random walk with Gnutella. We make statistics of query messages amount of each *TTL* round and the results are showed in Figure 7. It can be found significantly that as *TTL* increasing the amount of query messages of reliable random walk is not increasing sharply and is far less than that of Gnutella.



(a)τ=3.0,A=0.75,$Degree_{MAX}$=20 (b)τ=2.5,A=0.60,$Degree_{MAX}$=20

Figure 5. Query messages of Gnutella and reliable random walk

VI. SUMMARY AND FUTURE WORK

This paper presents reliable random walk, a resource search algorithm, whose query efficiency is scalable based on the scale-free feature of P2P network. Reliable random walk predicts the proportion of the hot nodes to quantify the resource search efficiency of the P2P network. Compared to other similar P2P network resource searching algorithms, reliable random walk is easy to be followed with low maintenance costs and need not store any additional information, and does not require statistical analysis on the information of the history search requests.

In future research work, we'll consider the introduction of local scale-free feature analysis to enable node's control of local P2P network's distribution so that query messages can be adjusted adaptively, which will make further improvement of the network query efficiency.

ACKNOWLEDGMENT

This research is partially supported by Self-research Program Foundation of Wuhan University under Grant 6082024, National Natural Science Foundation of China under Grant 70901060.

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] 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.

[6] 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), 2003, pp. 102-109.

[7] 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), 2002, pp. 84-95.

[8] 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.

[9] Rongmei Zhang, Y. Charlie Hu. Assisted Peer-to-Peer Search with Partial Indexing. In: Proc. of INFOCOM, 2005, 1514-1525.

[10] Abhishek Kumar, Jun Xu, Ellen W. Zegura. Efficient and Scalable Query Routing for Unstructured Peer-to-Peer Networks. In: Proc. of INFOCOM, 2005, 1162-1173.

[11] 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

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

[13] Nima Sarshar, P. Oscar Boykin, et al. Percolation Search in Power Law Networks: Making Unstructured Peer-to-Peer Networks Scalable. In: Proc. of the Peer-to-Peer Computing(P2P), 2004, pp. 2-9.

[14] 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.

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

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