Preferential Walk: Towards Efficient and Scalable Search in Unstructured Peer-to-Peer Networks

Hai Zhuge, Xue Chen and Xiaoping Sun
China Knowledge Grid Research Group, Key Lab of Intelligent Information Processing
Institute of Computing Technology, Chinese Academy of Sciences
Beijing, 100080, China
Zhuge@ict.ac.cn

ABSTRACT

To improve search efficiency and reduce unnecessary traffic in Peer-to-Peer (P2P) networks, this paper proposes a trust-based probabilistic search algorithm, called *preferential walk* (*P-Walk*). Every peer ranks its neighbors according to searching experience. The highly ranked neighbors have higher probabilities to be queried. Simulation results show that *P-Walk* is not only efficient, but also robust against malicious behaviors. Furthermore, we measure peers' rank distribution and draw implications.

Categories and Subject Descriptors: G.3

[**Mathematics of Computing**]: Probability and Statistics – Experimental design, Probabilistic algorithms.

General Terms: Algorithms, Design, Experimentation.

Keywords

P2P, search, probability, trust, power-law.

1. INTRODUCTION

The fundamental challenge in Peer-to-Peer (P2P) computing is how to provide efficient and scalable search services in a large-scale, open and dynamic environment. However, current search techniques used in existing P2P systems are not highly efficient or scalable. For example, the flooding search causes an overwhelming amount of traffic, which makes the Gnutella being far from scalable. The random walk can reduce traffic volume by forwarding queries to only a subset of neighbors [4]. But it cannot achieve high search efficiency due to the randomness nature in choosing neighbors.

In reality, peers are very heterogeneous in capability (storage, processing power, and bandwidth, etc), availability (the quality of being present or ready for immediate use [3]), and reliability (malicious peers always exist, and they can misbehave in many ways [5]). Thus, in an attempt to incorporate peers' heterogeneity into the design of search mechanism, we propose *preferential walk (P-Walk)*, a trust-based probabilistic search technique in an unstructured P2P network. A unique characteristic of our design is the utilization of a trust evaluation method to rate neighbors according to the feedback from previous searches. Neighboring peers assign each other trust ranks. During routing, peers preferentially forward queries to the highly ranked neighbors.

This work was supported by National Science Foundation (60273020 and 70271007) and National Basic Research Program of China (973 no.2003CB317001).

WWW 2005, May 10-14, 2005, Chiba, Japan.
ACM 1-59593-051-5/05/0005.

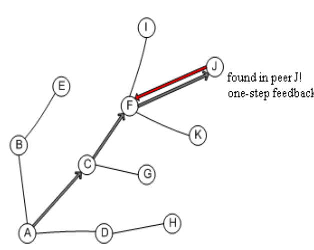
2. DESIGN OF P-WALK

In *P-Walk*, neighboring peers assign each other trust ranks. The trust rank for peer *B* at its neighbor *A* is a measure of how likely peer *A* considers that peer *B* will return the satisfactory resources when forwarding a query to *B*. The rank value is initialized to zero, and gradually updated based on *one-step feedback* mechanism, namely, when node *B* replies *A* with a *QueryHit* message, after validation, *A* will add one to peer *B*'s rank value.

Suppose peer *A* has *s* neighbors $N_1, N_2, \ldots, N_s$ ($s \geqslant 0$), and r_i indicates the rank peer *A* assigns to its neighbor N_i. Then, peer *A* will choose neighbor N_i as its query "receiver" with the forwarding probability:

$$p_i = \frac{(r_i + r)}{\sum_{n=1}^{s}(r_n + r)}$$

The risk factor *r* must satisfy the constraint $r > -r_i$ ($i = 1,2,\ldots, s$) to ensure that the forwarding probabilities are positive for all neighbors. The non-zero risk factor reflects some instances when a peer's rank cannot truly reflect its behaviors. For example, a malicious node may be in disguise and behave as a good one in the beginning, and other peers thus regard it as a good neighbor and assign it a relatively high rank. When $r = 0$, the forwarding probability for each neighbor is in direct proportion to its rank. Note that $\sum_{n=1}^{s}(r_n + r)$ converts this rate into a normalized probability, thus, $p_1 + p_2 + \ldots + p_s = 1$. Every neighbor has a certain probability to be the receiver. The higher the rank, the higher the probability.



direction	current rank	current probability	updated probability	updated rank
A to B	0	1/3	1/3	0
A to C	0	1/3	1/3	0
A to D	0	1/3	1/3	0
C to F	0	1/2	1/2	0
C to G	0	1/2	1/2	0
F to I	0	1/3	1/4	0
F to J	0	1/3	1/2	1
F to K	0	1/3	1/4	0

Figure 1. P-Walk procedure from peer *A* to peer *J*

Figure 1 explains the *P-walk* procedure. Query forwarding path follows the direction of the bold black arrow line until finding corresponding resource at peer *J*, and the red arrowhead line denotes one-step feedback from *J* to its predecessor *F*. Changes of the ranks and probabilities for each peer are summarized in the right chart.

3. EXPERIMENTAL EVALUATION

We design a Gnutella simulator based on the *Barabási-Albert* model with 1000 fixed peers [2], and the generated topology follows a two-segment power-law degree distribution[1] similar to the observed trend in [6]. Each peer generates 200 queries and issues one query per time slot. Peers may be good or malicious in the network and malicious behaviors are designed based on [5].

We also simulate three other related algorithms for comparison:

(1) Random walk (*R-Walk*): a peer forwards queries to a randomly chosen neighbor at each hop [4].

(2) Max-Degree-biased walk (*D-Walk*): a peer forwards queries to the highest-degree neighbor at each hop [1].

(3) Max-Feedback-biased walk (*F-Walk*): a peer forwards queries to the highest-feedback neighbor at each hop [7].

(4) Preferential walk (*P-walk*): a peer ranks each of its neighbors, and then preferentially forwards queries to the highly ranked neighbor with a high probability at each hop.

3.1 Simulation Results

Figure 2(a) shows the comparisons for four algorithms' success rates[2] as TTL varies. Figure 2(b) shows a metric for query response time — the number of queries that have successfully found resources at each hop.

The findings drawn from the simulations are as follows.

- *P-wa*lk achieves a very high query success rate and a fast response time. Peers using P-walk to forward queries are most likely to find the desired resources.
- *P-walk* is the most robust. The searching processes are guided by the ranks peers assigned to each other. *P-walk* minimizes malicious behaviors by reducing the messages forwarded to lowly ranked neighbors.
- *P-walk* exhibits the self-learning ability, which ensures that each peer's rank on its neighbors is always up-to-date. Thus, the latterly issued queries consume clearly shorter hops to find resources than the formerly issued one, which reduces large volumes of query and reply traffic.

3.2 Rank Analysis

The approach assigns a collective rank for each peer by averaging all the ranks given by its neighbors. Then, a collective rank distribution is defined as the number of peers whose collective ranks fall in the same intervals. The inset of Figure 3(a) illustrates an approximately linear relationship between each peer's collective rank and its degree. Generally, the more the degree, the higher the rank, and high degree peers are in the minority. Figure 3(b) shows the collective rank distribution exhibits a power-law-like trend when plotted on a double-logarithmic scale. These observations accord with what we expect. In real systems, there are only a small number of high-degree peers with server-like characteristics but large amount of peers are client-like [6].

[1] The degree distribution is a statistical property for a network topology denoting the number of peers that have i (i = 0, 1, 2, 3, …) open connections.

[2] Success rate is a ratio of the number of successfully completed queries to the total queries.

4. CONCLUSION

This paper presents the design and evaluation of *P-walk*, a trust-based probabilistic search method in a Gnutella-like P2P networks. Comparisons show that *P-walk* achieves a high query success rate and a fast response time while greatly reduces traffic volumes and alleviates malicious behaviors. The simplicity and scalability of *P-walk* make it feasible to be applied in large-scale systems [8, 9,10].

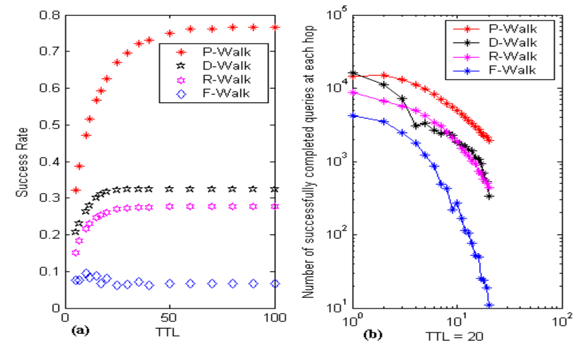


Figure 2. Performance comparisons

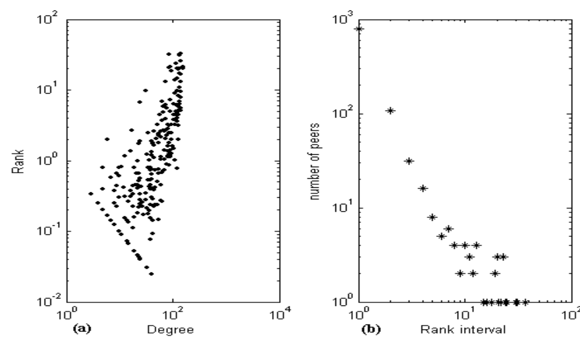


Figure 3. Collective rank distribution

5. ACKNOWLEDGMENTS

The authors thank all team members of China Knowledge Grid Research Group (http://kg.ict.ac.cn) and Knowledge Grid Center (http://www.knowledgegrid.net).

6. REFERENCES

[1] L.A. Adamic and R.M. Lukose, etc. Search in power-law networks, Phys.Rev.E, vol.64, pp.46135-46143, 2001.

[2] A.L.Barabási and R.Albert, Emergence of scaling in random networks, Science, vol.286, pp.509-512, 1999.

[3] R.Bhagwan, S.Savage and G.M. Voelker Understanding availability, IPTPS, 2003.

[4] C.Gkantsidis, M.Mihail and A.Saberi, Random walks in Peer-to-Peer networks, INFOCOM HongKong, China, 2004.

[5] M. Mishra, Cascade: an attack resistant peer-to-peer system, the 3rd New York Metro Area Networking Workshop, 2003.

[6] M.Ripeanu and I.Foster, Mapping the Gnutella network: macroscopic properties of large-scale peer-to-peer systems, IPTPS, Cambridge, MA, 2002.

[7] B.Yang and H.Garcia-Molina, Efficient search in peer-to-peer networks, ICDCS, Vienna, Austria, 2002.

[8] H.Zhuge, The Knowledge Grid, World Scientific Co., 2004.

[9] H.Zhuge, China's E-Science Knowledge Grid Environment, IEEE Intelligent Systems, vol.19, pp.13–17, 2004.

[10] H.Zhuge, The Semantic Grid: Scientific Issues, Architecture and Methodology. Communications of the ACM, Apr. 2005.