Notice of Retraction

After careful and considered review of the content of this paper by a duly constituted expert committee, this paper has been found to be in violation of IEEE's Publication Principles.

We hereby retract the content of this paper. Reasonable effort should be made to remove all past references to this paper.

The presenting author of this paper has the option to appeal this decision by contacting TPII@ieee.org.

2010 Third International Conference on Knowledge Discovery and Data Mining

Zipf's Trust Discovery in Structured P2P Network

Cai Biao
school of network
Chengdu University of Technology
Chengdu, P.R.China
caibiao @cdut.edu.cn

Chen Liangyin
School of Computer Science
Sichuan University
Chengdu, P.R.China
chenliangyin@gmail.com

Abstract*—*The use of peer-to-peer (P2P) applications is growing Dramatically. To finish transactions successfully, rust mechanism plays an important role, which not only compact the communication traffic but also the data discovery. In this paper we address the problem of trust discovery mechanism in structured P2P network we proposed before. The main contribution of this paper is address the Zipf's law to trust discovery. At last, the experimentally evaluate the effectiveness of uniform and zipf trust distribution, the result shows that Zipf's law performed advantages to uniform distribution.

Keywords-*p2p network;trust discovery;uniform distribution; Zipf's law distribution;structured toponogy*

I. Introduction

Peer-to-Peer network is a fully distributed computing model in which peers can directly communicate with others to exchange their information such as Gnutella[1], execute commercial transactions such as e-Bay[2] or watch video online such as PPLive[3]. In P2P communities, relationship between peers are often established dynamically and they are unrelated to and unknown to each other, so peers themselves have to manage the risk involved with the transactions without prior experience and knowledge about cooperators' reliability. One way to address this uncertainty problem is to make use of trust strategies to make peers can only interact with others based on their trust level.

Most existing trust models of e-commercial[4] require central entities for storing and distributing trust information of peers, so trust model for electronic markets cannot be directly transferred to a self-organizing P2P network. Peers in P2P network will play the same roles and there are no entities that can play a reliable trusted center for other peers will be essential. Then trust mechanisms in such network will be difficult and important. PeerTrust[5] model define a general trust metric based on three introduced basic trust parameters and two adaptive factors in computing trustworthiness of peers, which are feedback of a peers receives from others, the total transaction numbers a peer performed, the credibility of peers from feedback received, transaction context factor, and the network environment factor. EigenTrust[6] presents a distributed and secure method to compute global trust value, in which peers choose other peers from whom to interact with based on this global trust value and the trust model performed significantly to decrease the number of inauthentic files in P2P network. PowerTrust[7] aggregate global reputation with significantly and accuracy speed by using a look-ahead random walk strategy, and dynamically selects top-n power peers that are most reputable by using a distributed ranking mechanism. This trust model is robust to disturbance by malicious peers and adaptable to dynamics in peer joining and leaving. Our structured trust can give a accrue trust value manage scheme, in which trust information of everyone is fully entire distributed in network, and we proposed uniform hPSO method[8] based on fPSO[9] to fasten trust peer discovery implementation. Kacimi and Yetongnon[10] discussed *the similarity search in a Hybrid Overlay P2P Network in which organizes data and peers in a high dimensional feature space, all data and peers are described by a set of features and clustered using a density-based algorithm and evaluate the effectiveness of the similarity search with uniform and Zipf distribution.* In this paper we studied the similarity search in the Hybrid Overlay Network P2P. Learning-aware RPS[11] (LARPS) is proposed to overcome the disadvantages of high index cost of value index in DHT and high network traffic of recursive partition search (RPS) based on DHT in large scale network, experiments in that show the high efficient in large scale network in which data follows Zipf distribution. Hsueh et al.[12] investigated users' sharing behavior profile, and demonstrate the possibility to discover user characteristics in P2P environment by Zipf distribution. It is evident that any single document follows the long-tail distribution. Any resource corpus shared by users in a P2P environment also follows the long-tail distribution. Acosta and Chandra[13] showed that popular query terms remained stable over time exhibited a similarity of over 90% while there was little similarity over time (<20%) between popular file annotation terms and popular file terms, traditional P2P search focus on performance analysis did not take this mismatch between query terms and object annotations into account and thus overestimated performance of system. But trust value in our structured scheme is accure value and the mismatch isn't a consideration in the topology. So Zipf's law can be used to trust discovery in structured P2P network.

The rest of this paper is organized as follows: Section 2 will introduce the structured P2P network we proposed before and Section 3 is the details of Zipf distribution trust discovery. Section 5 is simulation performance results and the conclusion and future works is organized in section 6.

II. Our Structured Trust Topology

We have introduce the structured P2P this section is we proposed before that the critical technique for

978-0-7695-3923-2/10 $26.00 © 2010 IEEE
DOI 10.1109/WKDD.2010.93

structuralizing P2P network is the consistent hashing function, which assigns each peer an *m*-bit identifier using a general hash function such as SHA-1[14] in Pseudo Trust[15], and has features including the uniform distribution of outputs and resilience to collisions. We organize peers on DHT as Chord [16], and use the term “key” to refer the hashed value of the unique identifier under the hash function. Alls peer is organized in a circle according to their keys with increasing order.

To maintain trust management in such P2P network, every peer will assigned at least one trust manager to manage it’s identifier and trust information. The trust manager of peer *i* is assigned as follows: if peer *j* is a certain successor of h_i, where h_i is the hash value of its unique identifier of peer *i* hashed by the predefined hash function, then allocate the peer *j* as the trust manager to peer *i*. If other peers want to contact with a trustable peer *i*, they may issue a lookup on the manager to abstain the identifier information of peer *i* to get relate to the trustable peer. The key of the manager $h_{m(i)}$ is decided on a predefined constant *c* and calculated as $h_{m(i)}=(h_i+h_c)\mathrm{mod}2^n$, where h_c and $h_{m(i)}$ be hash values of constant *c* and identifier of the manager hashed by the predefined hash function respectively, and $h_{m(i)}=h_j$. If $m(i)$ isn’t in this circle, then $m(i)=m(m(i))$. PowerTrust[7] point out that multiple hash functions can be used to against malicious manager reports incorrect trust scores, here, we can adopt different constants instead of different hash functions to implement this purpose.

III. TRUST DISCOVERIES

Because peers in P2P network are dynamic and unknown to each other, so trust accumulating of a new peer random is a NPC complex problem. We have employed a developed history particle swarm optimization to implement a top-*n* trust accumulating in P2P network in which routing selection followed uniform distribution[8], while in this section, we will address the performance of peers which follow Zipf’s law based on hPSO.

A. History Particle Swarm Optimization

In this section, we introduce history Particle swarm optimization (hPSO) first. PSO is an evolutionary computation technology and was introduced by Eberhart and Kennedy[17] in 1995. Standard particle swarm optimization (SPSO) requires every particle in the swarm has two “best-positions” which are personal best position in history (p_0) and population best position be discovered so far (p_g). And particles will update their positions and velocity according to the two best positions to find the global optima iteratively. While in P2P network, when new p_g is find by one accumulating message, the p_g peer does not know where other accumulating messages will be, so other peers with messages arrived can not update their position according to the new p_g. One solution is to let peers who routing messages arrived can keep in attach with each other, but it obviously is illegitimate that it will produce much unnecessary traffic on communications for these relations. To overcome these challenges, we define a new history PSO (hPSO) algorithm based on fPSO [8] as follows:

$$x_j^{i+1}=c_1r_1(p_1-x_j^i)+c_2r_2(p_2-x_j^i)+\cdots+c_kr_k(p_k-x_j^i) \qquad (1)$$

where c_1 and c_2 are constants, *j* is the j^{th} dimension, x^i is position of i^{th} iteration, v^i is velocity of i^{th} iteration, p_l is the l^{th} best position of history of this particle in decreasing order, r_1, $r_2 \backsim U(0,1)$, and a parameter v_{Max} is used to constrain the moving scope of particles.

B. Uniform distribution Trust Discovery

However, the hPSO usually can be used to solve continuous problem, while peers in network are discrete contribution, so this equation cannot be directly transferred to P2P network for trust peer discovery procedure. Then we discrete the operational components of (1) with uniform distribution as follows [8]:

Positions: peers in P2P network x^i_j.

Subtractive operation between two positions: this operation is peers routing in P2P network, we mark peer *A* routing on peer *B* as $\Re = A\Theta B$.

Multiplicative operation between real number and peer routing: this operation is a routing decision operation when make a routing selection, we mark a real number *r* multiplies a routing decision $\Re(j)$ as $r\otimes\Re(j)$, where *r* is decided by a random number as follows:

$$\begin{cases} r=1 & \mathrm{int}((n+1)random)=j \\ r=0 & \mathrm{int}((n+1)random)\neq j \end{cases} \qquad (2)$$

where *n* is the number of trustable peers of top-*n*.

C. Zipf Distribution Trust Discovery

Due to the fast location of trust peer in structured trust topology, it is the most important issues for reducing address traffic, which accounts for a large amount of Internet traffic today. The less logic hop in location, the more efficient the address method will be. We have supposed the performance of uniform distribution during trust peer location, but researchers have observed that web pages are requested follows Zipf distribution[18]. Zipf's law states that the address probability of a request for the i'th most popular page in certain system is proportional to 1/i. The basic concept of Zipf’s law is[19,20] the address frequency of the s^th^ power of rank peer marked as *f*, and, rs, is a constant, *k.* Such relationship can be modeled by the *Zipf* distribution, which is also known as Pareto or long-tail distribution:

$$f\times r^s\approx k \qquad (3)$$

In this section, we will explore the applicability of Zipf law to trust peer discovery. As we discussed in section 1, trust information in our structured P2P network is accrue value, and then the content match will not be a consideration. To efficiently discover the trust peer for consumer, we use hPSO above mentioned to look forward method too, while the strategy in this section is Zipf instead of uniform distribution. The main difference between uniform and Zipf distribution is the multiplicative operation of routing selection. We express multiplicative operation in (4). As it showed in uniform distribution, the operation is a routing decision operation when make a routing selection, but the real number *r* is decided by Zipf distribution here.

$$\begin{cases} r=1 & v(d(i)))=f_i \\ r=0 & v(d(i)))\neq f_i \end{cases} \qquad (4)$$

where *n* is the number of trustable peers of top-*n*, fi is the access frequency of i^{th} popular peer, and the $v(d(i))$ of

i^{th} popular peer in N scale network will be calculated in next.

Then routing decision of trust peer discovery algorithm for a new peer in the network can be present as follows:

$$x_j^{i+1} = x^i + \sum_{l=1}^{k} r \otimes \Re(l) = x^i + \sum_{l=1}^{k} r \otimes (x_j^i \Theta p^l) \quad (5)$$

where x^i and x^{i+1} are current peer and the next peer which message will be send to respectively. In this top-n trust accumulating, the new peer will send trust query message to peers in its routing table, and every routing peer will make decision peer will be the next hop iteratively, until the top-n trustable peers have been found or remain hops equal to 0. Then, the top-n trust peer discovery algorithm for a trustable peer can be present as the follow algorithm:

```
Algorithm: top-n trusts discovery
Begin
Initial the trust vector of trustable peers
for every routing neighbor
do
send query message to routing neighbors
while (TTL>0 or condition isn't satisfied)
routing decision based on routing decision as (5)
else
return IP address of the trustable peers to requester
endwhile
enddo
End
```

D. *Real number* r *Calculating*

As showed above, the product of access frequency *f* and i^{th} peer is a constant k., and then Zipf's law can be expressed as:

$$n_i = k / i^{\alpha}$$

where n_i is the access number divided by network size N, then

$$\sum_{i=1}^{N} n_i = k \sum_{i=1}^{N} 1 / i^{\alpha} = 1$$

while peers exist in a large scale network, namely N is great large, then

$$k \cdot \int_0^1 1/t^{\alpha} dt = 1$$

where $t = i/N$, let $(t, t+\delta) \Leftrightarrow (i, i+1)$,then real number $v(d(i))$ can calculated as follows:

$$v(d(i)) = k\left(\int_0^{t+\delta} 1/t^{\alpha} dt - \int_0^{t} 1/t^{\alpha} dt\right)$$

In a certain network, N and α are constant, $\delta = 1/N$, t is decided by the order i^{th}, then if t is fixed, the real $v(d(i))$ can be conducted.

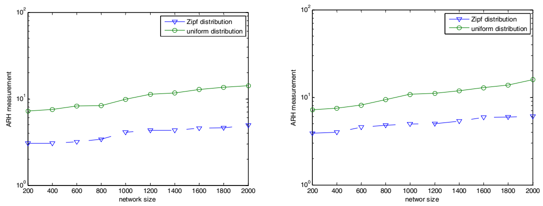
IV. Simulation Experiments

A. *Experiment Setting*

We simulated experiments on a simulation platform that we developed on Matlab 7.0. In next sections, trust score of peers is randomly produced and three trust level are given: (0.9, 1.0), (0.2, 0.3) and (0, 0.1) to express high trust, middle trust and low trust respectively, the portion of three levels in long tail distribution is 0.02, 0.88 and 0.1 respectively in the network. Because hash function can make peers in the circle balance arrangement, so we suppose key of peers are uniformly distributed with increasing one by one order on the DHT circle instead of arrangement of peers on the circle according to their hash value of identifier. Because this topology is focus on the comparison with uniform distribution, then we design experiment on the performance of effectiveness, traffic load of trust management when peers join in and peers leave out the structured circle.

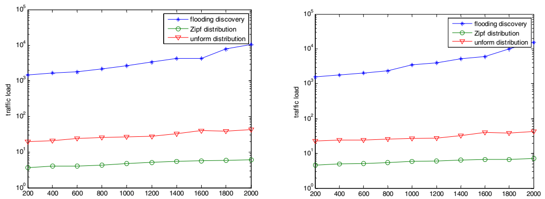
B. *Effectiveness of Trust Discovery*

We evaluate the effectiveness of trust discovery when issuing a trust query, and suppose there haven't malicious peers in the experiment. We compute the average path hops (APH) of 10 runs. The lower APH indicates the higher effectiveness of trust discovery performance. The APH is defined the number of logical hops between the trust trustor and trustee in the network. When peer get 5 peers whose trust score belong to (0.9, 1), we think the discovery procedure is finished. We plot the APH against different number of peers in a stationary network in Fig. 1(a), and in Fig. 1 (b) in a dynamic network with 10% peers join in or leave out. We set A random number implements the dynamics of 10%. When a routing message arrived on a certain peer, this peer will produce a random number r, if r>0.9, it will implement requirement of this message, otherwise, keep on muting. We find the results performance of uniform distribution is approximately to Chord, and there is obviously advantage that Zipf performed.



(a) in a stationary network (b) in a dynamic network

Figure 1: Effectiveness measure



(a) in a honest network (b) in a dishonest network

Figure 2: Traffic load measure

C. *Traffic Load Measure*

We evaluate the traffic load of trust discovery procedure, and suppose there isn't malicious peers and 20% malicious peers in the experiment respectively. We compute the average number of message (ANM) of 10 runs. The lower ANM indicate the smaller waste of resource in the procedure. The ANM is defined as the messages number of query fly in the network. Peers with trust score belong to (0, 0.1) will be isolated from the network. We plot the ANM against different number of peers in a stationary network in this simulation. Figure 4(a) shows the ANM of a trust discovery in an honest network and figure 4(b) in a dishonest network with 20%malicious peers. Malicious peers of 20% is decided by it trust score. If trust score *ts*>0.2, this peer is taken as an honest peer, otherwise, take it as a dishonest peer. The result also shows that the ANM of Zipf trust discovery in a dishonest network with 20% malicious peers performed advantage to uniform distribution, while the ANM of flooding scheme is about hundreds or thousands times to that of uniform and Zipf's law trust discovery procedures.

V. Conclusion

Based on the hPSO approach for trust discovery in structured P2P network, the uniform scheme for trust we have discussed before, but the Zipf long-tail distribution has been addressed for resource search last which can efficiently categorize the shared information with small portion of contents. In this paper, we employ the Zipf's law for trust discovery. The result of experiments shows Zipf's law performed obviously advantage than uniform distribution in trust discovery procedure. As for the future work, more investigation of trust discovery approaches is recommended. Another important issue is to take these schemes working on the e-commercial system we developed now.

References

[1] M.Ripeanu. Peer-to-peer Architecture Case Study: Gnutella, *In Proceedings of International Conference on P2P Computing' 2001*

[2] http://www.ebay.com

[3] http://www.pplive.com

[4] C. Dellarocas, The Digitization of Word of Mouth: Promise and Challenges of Online Reputation Mechanism, *Management Science*, vol. 49, no. 10, 2003.

[5] L. XIONG and L. LIU, "PeerTrust: Supporting Reputation-Based Trust for Peer-to-Peer Electronic Communities," *IEEE Trans. Knowledge and Data Eng*, vol. 16, no. 7, 2004, 843-857

[6] Sepandar D. Kamvar, Mario T. Schlosser, Hector Garcia Molina. The EigenTrust Algorithm for Reputation Management in P2P Networks. *In ACM proceedings of WWW2003*, *May* 2003.

[7] Runfang Zhou, Kai Hwang. PowerTrust: A Robust and Scalable Reputation System for Trusted Peer-to-Peer Computing, *IEEE Transactions on Parallel and Distributed System*, Vol.18,(4).2007, 460-473

[8] CAI Biao, Li Zhishu, Cheng Yang, FU Die, Cheng Liangyin. Trust Decision Making in Structured P2P Network, *In Proceedings of ICCSN2009*, pp679-683.

[9] CAI Biao, Li Zhishu, Fu Die, Hu Jian, Li Qing. Mutated Fast Convergent Particle Swarm Optimization and Convergence Analysis, *In Proceedings of ICINIS2008*, pp5-8.

[10] Mouna Kacimi, Kokou Y´etongnon. Density-based Clustering for Similarity Search in a P2P Network. *CCGrid'06*, pp57-64

[11] Ze Deng, Dan Feng, Ke Zhou, Zhan Shi, Chao Luo. Range Query Using Learning-Aware RPS in DHT-Based Peer-to-Peer Networks. *CCGrid'09*, pp180-187

[12] Hsiang-Yuan Hsueh, Jing-Shiuan Hua, Shi-Ming Huang, Hartmut J. Will. Resource Sharing Behavior in a Socialized Peer-to-Peer Internet Environment, *ICMB'09*, pp131-136

[13] William Acosta, Surendar Chandra. On the need for query-centric unstructured peer-to-peer overlays. *IPDPS'08* pp1-8

[14] FIPS 180-1. Secure Hash Standard. U.S. Department of Commerce/NIST, *National Technical Information Service*, Springfield, VA, Apr. 1995

[15] Li Lu, Jinsong Han, Yunhao Liu, Lei Hu, Jinpeng Huai, Lionel M. Ni and Jian Ma. Pseudo Trust: Zero-Knowledge Authentication in Anonymous P2Ps [J]. *IEEE Transactons on Parallel and Distributed Systems*, 2008,VOL. 19, NO. 10, OCT. pp1325-1337.

[16] Ion Stoica, Robert Morris, David Karger, et al. Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications, *In ACM proceedings of SIGCOMM'01*, Aug. 2001, 149-160

[17] Kennedy J, Eberhart R.C. Particle Swarm Optimization. In *Proceedings of the IEEE Conference on Neural Networks*, IV. 1995,1942-1948

[18] George Kingsley Zipf. Relative frequency as a determinant of phonetic change. Reprinted from the Harvard Studies in Classi cal Philiology, Volume XL, 1929.

[19] C. van Rijsbergen, Information retrieval, 2nd ed., London: Butterworths Publication, 1979.

[20] L. Adamic and B. Huberman, "Zipf's law and the Internet", Glottometrics, vol. 3, 2002, pp143-150.