Recent Advances in Research on P2P Networks

Yunfei ZHANG[1,2], Changjia CHEN [2], Xiaoping WANG[1]
[1](Department of Broadband Network, Shanghai Institute of China Telecom, Shanghai 200122, China)
[2](School of Electrics and Information Engineering, Beijing Jiaotong University, Beijing 100044, China)
hishigh@sohu.com

Abstract

As one of the killing applications in NGI, peer-to-peer networks (P2P for short) have rapidly developed in recent years. We survey and catalog the current hot research aspects in P2P networks, compare and review most of the research work in P2P networks, and summarize the research means and their problems.

1. Introduction

As one of the killing applications in NGI, peer-to-peer network (P2P for short) has rapidly developed in recent years. Although P2P is just popular in recent years, its conception has been proposed ever since 1960s. Both ARPANET [1] and Usenet [2] are of P2P characteristics, i.e., they are distributed and decentralized file transfer and sharing systems. According to different network structures, P2P systems can be divided into three generations:

Napster[3] belongs to the 1st generation P2P system (1GP), which includes server and P2P nodes. The server indexes files stored in P2P nodes and returns the owner nodes' information when receiving requests from the searching nodes. With the owner nodes' information, requesting nodes connect with the owner nodes directly to get the destination files. But because Napster is suspected to violates digital copyrights, it is accused by Recording Industry Association of America (RIAA) and finally closed. Audiogalaxy[20]and Scour[21]are also closed for the same reason. A recent news says that Napster is reopened by settling the digital copyrights problem. The biggest problem in the 1st generation P2P is its bad scalability because of its central indexing server, which also cause single point failure.

To overcome the problems in the 1st generation P2P, there are great changes in the 2nd generation P2P (2GP) that is based on distributed indexing. Gnutella [4] and Fasttrack[22]are typical 2GP. The former has totally unstructured topology and the latter has hierarchical topology. 2GP searches files based on flooding, which brings out useless traffic in Internet. Some other systems like Edonkey[6]/Emule[7] (with distributed indexing servers) and Bittorrent[23](with trackers to provide downloading nodes' information with the same requested files to the requesting nodes)are also of 2GP structurally.

The 3rd generation P2P (3GP) employs structured topology where P2P nodes array by their nodeIDs lexically. Meanwhile it builds mapping between fileIDs and nodeIDs and stores files in corresponding nodes according to the mapping. Here structured topology includes ring, Cartesian plane, supercube and tree topology. 3GP searches files based on Distributed Hash Table (DHT) routing. Because of the pre-built mapping between fileIDs and nodeIDs, DHT provides efficient routing and generally files can be found within $O(logN)$ steps where N is the number of P2P nodes. There are a lot of 3GP applications, such as Chord [24], CAN [25]and Pastry[26].

In the following paper we survey the main research aspects, solutions and key technologies in P2P networks, including topology building, routing optimization, key words partial matching search in DHT routing, semantic P2P, P2P network measurement, P2P performance analysis, P2P system modeling, incentives in P2P system and novel applications deploying P2P systems. We summarize this paper and discuss the future work in the last section.

2. Research on P2P Networks

Nowadays P2P network has been paid much attention in academic fields. Many algorithms to improve P2P performance have been proposed. We catalog hot topics in P2P research into the following aspects:

2.1. Topology Building of P2P Networks

Proceedings of the Seventh International Conference on
Parallel and Distributed Computing,Applications and Technologies (PDCAT'06)
0-7695-2736-1/06 $20.00 © 2006 IEEE

The general process to build P2P network topology is as follows: the joining node selects neighborhood node and their connections according to some certain rules to make P2P network tend to be some certain topology. This is a gradual process, which is different from 3GP. The latter sets a certain topology in advance. The means to build P2P topology include:

Building Selective Connection with connectivity information: R.H. Wouhaybi in [27] selects suitable nodes to build "selective connections" according to local connectivity information to gradually realize a Power-Law topology. In [28] L. Ramaswamy employs distributed clustering algorithm to build clustering topology, in which clustering nodes flood their connectivity and other nodes judge the weights of every clustering node which represent the degree of connectivity between clustering nodes and other nodes. A node selects the clustering node with the heaviest weight to build clustering topology.

Using Detector to update connectivity: In [29] Y. Liu makes topology matching in unstructured P2P systems by getting rid of low efficient connections and building near connections in topology. Every node runs a detector and send detective information to neighborhood nodes periodically (TTL=1 and 2); other nodes receive detective information from different paths and calculate delay of each path. According to the delay, redundant and low efficient links are got rid of and nearer connections are built. The experiments show that this solution is effective to reduce 75% communications traffic and 65% response time.

Game theory: B. Chun [30] changes topology building into game theory problem: On one side, a node want to maximize its connections amount; but on the other side, a node wouldn't share too much with other nodes. The paper builds a cost function by connection expense and nodes distance and every node runs adding-link and deleting-link algorithms periodically to change the connectivity of P2P systems to minimize the total cost. Different cost function affects the topology much and node degree varies from exponential distribution to power-law distribution.

2.2. P2P Routing Optimization

Today P2P routing optimizations include two parts: one is routing optimization in unstructured P2P systems and the other is routing optimization in structured P2P systems.

Routing optimization in unstructured P2P systems solves the problem of flooding, which brings redundant offsets and traffic. The solutions have:

Random Routing (Walk): Also called probabilistic routing, which means a node no longer floods the requests, but selects one or more next-step nodes by probability randomly or with certain rules. In case of not decreasing search scale, the request traffic decreases by two orders of magnitude, but the response time increases by one order of magnitude [31]. The strength of random routing is that its routing termination condition is self-adaptive. C. Gkantsidis in [32] views random routing as a Markov process and finds sampling of consecutive two steps in random routing is same as independent sampling statistically.

Improvement in Random Routing: The biggest problem in random routing is that the routing is ineffective if the next-step node has little forwarding information and there will be a long queue time before forwarding if the next-step node is of high load [33]. To solve this problem, Adamic, L. A et.al selects nodes with high connectivity. The intuition behind this is that nodes with high connectivity have more connectivity information [34]. D. Tsoumakos uses history record as the guide in next step searching and acquires 3 times more results than random routing in case of same bandwidth consumption [35].

QoS Routing: Zhu proposes QoS routing problem in unstructured P2P networks [36]. Suppose every node can get its own global sorting rank according to some QoS evaluation, the forwarding rule is: forwarding the request to the highest QoS neighborhood node if there exist neighborhood nods with higher QoS level.

DHT routing optimization focuses on how to find physically closer nodes in the next-step forwarding. The solutions include:

Updating Routing Table by Sampling: H. Zhang [37] uses DHT routing to measure the delay from one node to some other candidates and uses sampling to select candidates with lower delay to update DHT routing table to decrease searching delay incrementally. In Proximity Neighborhood Selection (PNS), K. Gummadi classifies proximity into two types [38], in which PNS(K) samples in the consecutive K nodes after one node in a routing interval (e.g., Chord) randomly and selects the physically closest nodes as the forwarding node in this step. By investigating routing delay distribution, we can get suitable K.

Adding New Data Structure in Routing Table to change searching way: Pastry[39] employs neighborhood set to keep physically near nodes record and compare logical neighborhood nodes in routing table and physical neighborhood nodes in neighborhood set to find suitable physical neighborhood nodes replacing current routing table items. This will ensure that every step is closest in possible routings. In P2P systems the ratio of logical one-step delay and physical one-step delay is called *routing stretch*, which is a measure to evaluate the routing efficiency. The smaller routing stretch is, the more efficient the routing is. Rhea [40] finds that the large routing stretch often occurs in where source node and destination node are of near distance. To solve this, in Tapstry the author proposes a novel routing algorithm based on a special data structure called "Attenuated Bloom Filter". Like Pastry, a node uses attenuated bloom filter to record data stored in all physical neighborhood nodes within *D* steps. For each request, the nodes checks its attenuated bloom filter to find the matching information; if it doesn't exist, the node forwards the request using DHT routing.

Changing Node Join Process: Ratnasamy[41] matches physical network and application network by a distributed binning scheme in CAN. The binning scheme sets up *k* evenly distributed nods as landmark nodes in the Internet and the key space of CAN is a *d*-dimensional Cartesian plane. The key space is divided into *k!* equal sized portions, each corresponding to a single ordering of k-dimensional coordinates. A new node joins its portion according to its distance to landmark nodes which makes close nodes gather into near portions. M.Waldvogel proposes a similar algorithm as binning scheme in Mithos [42]. Differently, a new node calculates its nodeID according to its neighborhood nodeID and the delay between the new node and the neighborhood node. The finding of nearest neighborhood node is similar as Pastry [39]. In routing, a node builds connectivity with its nearest neighborhood node in every dimension and forwards requests to the neighborhood set in the same dimension.

2.3. Key Words Partial Matching Search in DHT Routing

DHT only supports precise content searching, i.e., a search can be done only when we know the exact name of the target. Two files with similar names can be stored in quite distant nodes in 3GP, which is contraventional to our daily key words searching. To solve this, key words partial matching search is proposed in DHT routing. The basic idea is to build indexing from key words to fields and according to this indexing, find fileID in key word partial matching search; then do precise content searching with the fileID [43][44][45][46]. To be concrete, building possible partial key words set $Q=\{q_1, q_2, \ldots, q_l\}$ for file *f*, and according to DHT rules, keeping index of (q_i, Q) in the node whose nodeID is $hash(q_i)$. When searching file *f* , using partial key word q_i to find the index of (q_i, Q). Then according to this index, finding the exact fileID *f*, then using DHT routing to find file *f* eventually. In this way, search of any key word in *Q* will contribute to find file *f*.

Overnet[47] is the only commercial software to support key words partial matching search. But what's different, the key word indexing stores in different position. Every node keeps some key words indexing lexically and files corresponding to the key words are stored according to DHT rules. Overnet separates coarse request of file *f* into requests of some independent key words and nodes storing corresponding indexing return the storing nodes of these key words. The requesting node then sends requests to these storing nodes to get the target.

2.4. Semantic P2P

The emergence of more and more P2P systems brings the problem s of searching efficiency and interoperation. There are many kinds of network devices which are not easily compatible one another. In addition, the variation information formats, polysemy and shortage of information connections are very serious. How to utilize network resource to search information effectively is the main problem semantic P2P faces with. Traditional P2P doesn't support semantic forwarding (e.g., using Metadata) and this is mainly because t*raditional P2P lacks useful global information such as content distribution information and the relationship between key words and contents*. Gnutella is such an example: there is much useless information, which is caused by unorderly organized content stored in the network; meanwhile searching mechanism in Gnutella only supports simple text matching while doesn't support semantic searching based on content correlation.

The basic idea of semantic P2P is to introduce "Semantic" conception in traditional P2P and organize information resource by semantic rules to make better formalizing description. Based on this, the information in P2P is cataloged and redundant information is got rid of. What's more, semantic P2P employs *content routing* strategy; a node specifies "Predication" as requesting/forwarding content filter rules and every forwarding node builds an updated *knowledge base* by self-study and forwards incoming request based on knowledge base. Experiments of semantic P2P applied in Gnutella show that it saves 80% traffic in searching while keeps the amount of searching results.

In semantic P2P, there are several ways to classify content:

RDF Classification: The papers [48][49][50] use RDF（Resource Description Framework） in W3C to define metadata and search in P2P networks, which provides more explanatory and reasoning ability in reading data. But RDF is coarse granule content classification mainly used in Web service, which is much different from the requirement of clustering by semantic in P2P file sharing systems.

Global File Type Distribution Chart Classification: In SON[51], a node collects file type information by known classification standard to form global file type distribution view and put itself some certain content cluster. In this way semantic clustering is realized. But SON only focuses on well-classified file systems such as music files and it doesn't relate to more complex semantic classification.

Vector Space Model Classification: Deriving from information retrieval (IR), the papers [52] [53] [54] [55][56] investigates Vector Space Model (VSM) and builds *Inverse Indexing* for every key word to cluster contents. This is a good and file clustering solution used widely in recent years. In this model, files and requests are both represented by a vector space corresponding to a suit of orthogonal term vectors called *Term Vector*. Each element in the vector corresponds to the importance of a term in the files or requests. The importance of a term is determined by the frequency occurred in its own flies (term frequency) and its frequency occurred in other files (reverse file frequency). The more frequent a term occurs in it own file, the more representative it represents the file; while the more frequent a term occurs in other files, the less representative it represents its own file. In 3GP a file is stored with normalized eigenvectors of the key words as the key in a node. In this way semantic similar files are gathered. In searching, we use eigenvectors of the key words as the key, and search target files in a small region centered at the key. The node returns documents with high relevance according to the similarity of query vector and document vector. The drawback of VSM is its high cost of vector presentations in case of P2P churns. Also, the storage of popular key words indexing needs cache mechanism to support [53].

C. Tang [52] projects high dimensional representation of documents in VSM into low dimensional *Latent Semantic Space* through *Singular Value Decomposition* (SVD) to build *Latent Semantic Index* (LSI), which can reduce searching scale and avoids data sparseness.

Building relational database for key words: N. Ejdl. [57] introduces relational database for key words in case when there is no uniform data structure. Every node records relations among local key words and starts *Query Expansion* in unstructured P2P systems; every forwarding node expanses the relations between query and local database to globalize its key words database. This solution is suitable for multimedia data search with comparatively less key words.

2.5. P2P Network Measurement

Network measurement is the hot topic recently. It is trying to improve network performance through studying traffic and network behavior. We took the 1st P2P traffic measurement and analysis in Internet backbone in China in 2002 [58]. Nowadays P2P traffic accounts more and more in Internet [59].

2.5.1. Dynamics of P2P Networks. The dynamics of P2P networks usually employs active measurement to validate P2P node transit.

Connection Duration: S. Sen finds in[60]that more than 20% nodes keep connections less than 1 min, and 60% IP addresses keep connections less than 10 mins. S. Saroiu shows the average online time in Gnutella and Napster is 60 mins and online nodes send 0.3 queries per minute [61]. J. Chu in [62] measures the availability of some thousand Gnutella nodes and finds 31% nodes run 10 mins a day. Note that the three measurements are almost conducted in the same time and their results are comparable.

Node Join and leave: R. Bhagwan measures the availability of 2,400 Overnet nodes in Jan. 2003 and finds that a node joins and leaves P2P systems 6.4 times a day and over 20% nodes join and leave P2P systems every day.

All these research point out that transit of P2P nodes, which should be considered in P2P system design, e.g., Y. Liu builds P2P model with the above characteristics in validating topology matching algorithm.

2.5.2.Traffic Measurement in P2P Networks. Active measurement and passive measurement are both widely used in P2P traffic measurement. K. P. Gummadi et al. [65] measures Kazaa traffic in Univ. of Washington by active measurement. The measurement lasts for 200 days but note that most of the users are of free large bandwidth rather than common P2P users, the result is not well applied in other cases. S. Sen employs passive measurement to measure the popularity of Kazaa, DC and Gnutella and makes elaborate analysis for an ISP backbone in AT&T[66].

2.5.3. Measurement on file download performance. In the 1st research on P2P file download performance, B. Huberman et. al uses 35,000 Gnutella nodes in a day as the sample and finds that 70% users don't share any bandwidth to become "free-rider"[67]. In [68] S. Saroiu measures the bandwidth of 223,000 Gnutella nodes using SProbe. He finds that about 8% nodes download files at a speed less than 8KB/s.

[69] is the only research work to study content publishing time, which finds 77% unauthorized movie copies are leaked by industry insiders. What's worse, most movies appear on file sharing networks prior to their official consumer DVD release date.

B. Cohen makes preliminary measurement and studies Flash Crowd Effect in Bittorrent [70]. This effect is "created" by some thousand of volunteers downloading a test document asked by the father of Bittorrent, B. Cohen. N. Leibowitz studies the variation of document popularity within lifetime in Kazaa[71] and he doesn't find flash crowd effect. The possible reason is that unlike Bittorrent, Kazaa doesn't mainly share newest contents.

J.A. Pouwelse [72] takes the longest P2P measurement started from June, 2003. It lasts for 8 months. The author measures publishing websites and their mirrors in Bittorrent systems, .Torrent file servers, trackers and Bittorrent clients, studies popularity, availability, download speed, content lifetime and pollution level, where pollution level is first proposed. File pollution in P2P means transferring false content to pollute genuine files. Overpeer Co. [73] claims that it is polluting P2P networks by false MP3 contents to prevent copyrights of more than 30,000 names. J.A. Pouwelse [72] measures content lifetime in P2P systems and points out that if only one node keeps online, the content can survive in Bittorrent and there is little relations between file availability and the number of seeds.

2.6. P2P Performance Analysis

2.6.1. Availability. G. On takes availability as the metric of network QoS and proposes the conception of QoA, which is the ratio of the availability the user requires and the availability the system can provide [74]. Suppose different nodes have different QoA requirements, the paper simulates P2P system as dynamic random graph and investigates content placement in P2P networks according to QoA metrics. The experiments show that compared with the number of content copies, the placement of content is more important to QoA.

2.6.2. Scalability. Gia [75] discusses the scalability of Gnutella and increases the capacity 3 to 5 times. In Gia there are four measures to improve the scalability of Gnutella: First, it deploys dynamic topology adoption protocol to make most of nodes are adjacent to high capacity nodes with higher degree; second, it uses active flow control protocol based on token to avoid node overloading; third, it builds one-step content pointer to ensure every node keep pointers to the contents in its adjacent nodes. This is in accordance with the idea of topology improvement; Last, it employs selective random walks routing protocol, which uses the result of routing optimization in subsection 2.2.

2.6.3. Resiliency. The resiliency of P2P systems is closely related to P2P topology and the main techniques include:

Probing: R.H. Wouhaybi realizes P2P resiliency by two means [76]: One is to identify vicious nodes and hide the nodes with high connectivity for these vicious nodes. This can avoid vicious nodes attacking P2P systems. Here vicious nodes are nodes trying to capture global network topology by probing node connectivity. The other means is to probe nodes state to get rid of invalidated nodes when the system failure occurs.

Graph Theory: L. Nowell in [77] investigates Chord resiliency in the case of node joins and leaves and deduces the probabilistically lower bound of connectivity degree of nodes in a connectivity graph. K.P. Gummadi finds that the routing is more flexible for ring topology [78], and it can provide better performance in node failure. D. Loguinov applies optimal-diameter de Bruijn graph in P2P system [79] and studies routing, graph expansion and clustering and path overlapping. He points that optimal-diameter de Bruijn graph has the highest node connectivity among existing schemes. He also puts forward how to build de Bruijn graph.

3. Summary

In this paper we survey and catalog the current hot research aspects in P2P networks, compare and review most of the research work in P2P networks, and summarize the research means and their problems. For the limit of space, some important research aspects such as P2P system modeling, incentives in P2P systems and novel applications deploying P2P systems are omitted here. Interested readers can refer to [80] for more details.

4. References

[1].http://www.dei.isep.ipp.pt/docs/arpa.html
[2].http://www.usenet.com
[3]. http://www.napster.com/
[4]. http://www.gnutella.com/
[5].http://www.kazaa.com
[6]. http:// www.edonkey2000.com
[7]. http:// www.emule-project.net
[8]. http:// bitconjurer.org/BitTorrent
[9]. http://www.entropia.com
[10]. http://www.openp2p.com/pub/d/241
[11]. http://setiathome.ssl.berkeley.edu
[12]. http:// www.ud.com
[13]. http://www.msn.com
[14]. http://www.icq.com
[15]. http://www.jabber.org
[16]. http://www.tencent.com
[17]. http://www.skype.com
[18]. http://www.groove.net
[19]. http://www.zaplet.com
[20]. http://.www.audiogalaxy.com
[21]. http://www.Scour.com
[22]. http://www.fasttrackdcn.net/
[23]. http:// bitconjurer.org/BitTorrent

[24].Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, Hari Balakrishnan, Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications, In *Proc. ACM,SIGCOMM,*pp. 149-160, San Diego, CA,August 2001.

[25].Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp, Scott Shenker, A Scalable Content-Addressable Network, In *Proc. ACM,SIGCOMM* , pp. 161-172,San Diego, CA, August 2001.

[26]. Antony Rowstron1 and Peter Druschel, Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems, in *Proc. of the 18th IFIP/ACM International Conference on Distributed Systems Platforms*, pp.329-350, Heidelberg, Germany, November 2001.

[27]. Rita H. Wouhaybi, and Andrew T. Campbell，Phenix: Supporting Resilient Low-Diameter Peer-to-Peer Topologies，*infocom'04*

[28]. Lakshmish Ramaswamy, Bugra Gedik and Ling Liu,Connectivity Based Node Clustering in Decentralized Peer-to-Peer Networks, *Third International Conference on Peer-to-Peer Computing*,2003. Linköping, Sweden. p.66

[29]. Yunhao Liu, Xiaomei Liu,Li Xiao ,Location-Aware Topology Matching in P2P Systems, IEEE INFOCOM 2004

[30]. Byung-Gon Chun,Rodrigo Fonseca, Ion Stoica, and John Kubiatowicz ,Characterizing Selfishly Constructed Overlay Routing Networks, *Infocom'04.*

[31]. Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, "Search and replication in unstructured peer-to-peer networks," in *International Conference on Supercomputing (ICS'02)*. ACM, 2002.

[32]. Christos Gkantsidis, Milena Mihail, and Amin Saberi ,Random Walks in Peer-to-Peer Networks, *infocom'04*

[33]. LV, Q., Ratnasamy and Shenker, S. Can Heterogeneity Make Gnutella Scalable. In *Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02)*, MA, Mar. 2002.

[34]. Adamic, L. A., Lukose, R. M., Puniyani, A. R., AND Huberman, B. A. Search in Power-law Networks. *Physical Review E 64* 2001.

[35]. DimitriosTsoumakos, Nick Roussopoulos ,Adaptive Probabilistic Search for Peer-to-Peer Networks，*P2P'0*3

[36]. Cheng Zhu, Zhong Liu, Weiming Zhang, Weidong Xiao, Dongsheng Yang ,Analysis on Greedy-Search based Service Location in P2P Service Grid，*P2P'03*

[37]. Hui zhang.Incrementally Improving Lookup Latency in Distributed Hash Table Systems,*SIGMETRICS'03*

[38]. Krishna Gummadi,Ramakrishna Gummadi,. Sylvia Ratnasamy,The Impact of DHT Routing Geometry on Resiliency and Proximity.,*signalcom'03.*

[39]. C.G. Plaxton, R. Rajaraman, and A.W Richa.,Accessing Nearby Copies of Replicated Objects in a Distributed Environment. *the Ninth Annual ACM Symposium on Parallel Algorithms and Architectures,* 1997.

[40]. S. C. Rhea, J. Kubiatowicz. Probabilistic Location and Routing. *IEEE INFOCOM*, 2002.

[41].S. Ratnasamy, M. Handley, R. Karp, and S. Shenker. Topologically-Aware Overlay Construction and Server Selection. *IEEE INFOCOM*, 2002.

[42].M.Waldvogel, R. Rinaldi. Efficient Topology-Aware Overlay Network. *First Workshop on Hot Topics in Networks (HotNets-I)*.October 2002

[43]. M. Harren, J. Hellerstein, R. Huebsch, B. Loo, S. Shenker, and I. Stoica. Complex Queries in DHT-based Peer-to-Peer Networks. *Proc. IPTPS 2002*, Mar. 2002.

[44]. C. Tang, Z. Xu, and S. Dwarkadas. Peer-to-Peer Information Retrieval Using Self-Organizing Semantic Overlay Networks. *Proc. ACM SIGCOMM 2003*, Aug. 2003.

[45]. P. Reynolds and A. Vahdat. Efficient Peer-to-Peer Keyword Searching. *Proc. Middleware2003*, June 2003.

[46]. Matthew Harren, Joseph M. Hellerstein, Ryan Huebsch, Boon T. Loo, Scott Shenker, and Ion Stoica. Complex queries in dht-based peer-to-peer networks. *IPTPS02*, Cambridge, MA, March 2002.

[47]. http://www.zeropaid.com/bbs/archive/index.php/t-6928

[48].W. Nejdl, B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmer,and T. Risch. EDUTELLA:A P2P Networking Infrastructure Based on RDF. *Proc.WWW 2002*, May 2002.

[49]. J. Broekstra, M. Ehrig, P. Haase, F. Harmelen, A. Kampman, M. Sabou, R. Siebes,S. Staab, H. Stuckenschmidt, and C. Tempich. A Metadata Model for Semantics-Based Peer-to-Peer Systems. *Proc. SemPGRID'03*, May 2003.
[50]. A. Castano, S. Ferrara, S. Montanelli, E. Pagani, and G.P. Rossi. Ontology-Addressable Contents in P2P Networks. *Proc. SemPGRID'03,* May 2003.
[51]. Arturo Crespo and Hector Garcia-Molina ,Semantic Overlay Networks for P2P Systems, *Technical. report*, Stanford University, January 2003.
[52]. Chunqiang Tang.Zhichen XuSandhya Dwarkadas, Peer-to-Peer Information Retrieval Using Self-Organizing Semantic Overlay Networks, *ACM sigcomm03.*
[53]. LV, Q., Ratnasamy and Shenker, S. Can Heterogeneity Make Gnutella Scalable. In *Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02)*, MA, Mar. 2002.
[54]. LI, J., Loo, B. T., Hellerstein, J., Kaashoek, F., Karger,D. R., AND Morris, R. On the Feasibility of Peer-to-Peer Web Indexing and Search. In *Proceedings of the 2nd International Workshop on Peer-to-Peer Systems(IPTPS '03).* Berkeley, CA, Feb.2003
[55]. Reynolds, P., AND Vahdat, A Efficient Peer-to-Peer Keyword Searching. *Technical report,* Duke University, Durham, NC, 2002. Available at http://issg.cs.duke.edu/search/.
[56].Tang, C., Xu, Z., AND Mahalingam, M. pSearch: Information Retrieval in Structured Overlays. In *Proceedings of the ACM HotNets-I Workshop* (Princeton, NJ, Oct. 2002).
[57]. N. Ejdl. , W., M. W. Olpers. , W. S. Iberski, Semantic Peer-to-Peer Search Using Query Expansion，*Proc. SemPGRID'04*
[58].Yunfei Zhang, Lianhong Lei and Changjia Chen,Characterizing Peer-to-Peer Traffic across Internet，*Lecture Notes in Computer Science*, vol.3032, pp.388-395,
[59].http://it.sohu.com/20040925/n222233250.shtml
[60]. S. Sen and J. Wang, "Analyzing peer-to-peer traffic across large networks,"in *Proceedings of ACM SIGCOMM Internet MeasurementWorkshop*, 2002.
[61]. S. Saroiu, P. Gummadi, and S. Gribble, "A Measurement Study of Peer-to-Peer File Sharing Systems," in *Proceedings of Multimedia Computing and Networking (MMCN)*, 2002.
[62.] J. Chu, K. Labonte, and B. Levine, "Availability and locality measurements of peer-to-peer file systems", *ITCom: Scalability and Traffic Control in IP Networks*, July 2002.
[63]. R. Bhagwan, S. Savage, and G. M. Voelker, "Understanding Availability,"in *Proceedings of the 2nd International Workshop on Peer-to-Peer Systems(IPTPS'03)*, 2003.
[64]. Yunhao Liu, Xiaomei Liu,Li Xiao ,Location-Aware Topology Matching in P2P Systems, IEEE INFOCOM 2004
[65]. K. P. Gummadi et al., "Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload", *19th ACM Symposium on Operating Systems Principles(SOSP'03)*, Oct. 2003.
[66]. S. Sen and J. Wang, "Analyzing Peer-to-Peer Traffic Across Large Networks", *ACM/IEEE Transactions on Networking*, Vol. 12, No. 2, April 2004, pp. 137–150.
[67]. Adar, B. Huberman, ``Free riding on Gnutella'', *Technical report*, Xerox PARC, 10 Aug. 2000.
[68]. S. Saroiu, P. K. Gummadi, S. D. Gribble, ``A Measurement Study of Peer-to-Peer File Sharing Systems'', *Multimedia Computing and Networking 2002 (MMCN '02)*.
[69]. S. Byers, L. Cranor, E. Cronin, D. Kormann, P. McDaniel, ``Analysis of Security Vulnerabilities in the Movie Production and Distribution Process", *the 2003 ACM Workshop on DRM*, Oct. 2003.
[70]. B. Cohen, Incentives Build Robustness in BitTorrent, *bitconjurer.org/BitTorrent*, May 2003.
[71].N. Leibowitz, M. Ripeanu, A. Wierzbicki, ``Deconstructing the Kazaa Network'', *3rd IEEE Workshop on Internet Applications (WIAPP'03)*, June 23-24, 2003, San Jose, CA.
[72]. J.A. Pouwelse, P. Garbacki, D.H.J. Epema, H.J. Sips, A Measurement Study of the BitTorrent Peer-to-Peer File-Sharing System, *technical report PDS-2004-003*, submitted for publication, April 2004
[73]. Nieman, Digital Decoys, *IEEE Spectrum*, May 2003, p. 27.
[74].Giwon On, Jens Schmitt, Ralf Steinmetz,The Effectiveness of Realistic Replication Strategies on Quality of Availability for Peer-to-Peer Systems,*P2P'03.*
[75]. LV, Q., Ratnasamy and Shenker, S. Can Heterogeneity Make Gnutella Scalable. In *Proceedings of the 1stInternational Workshop on Peer-to-Peer Systems (IPTPS '02)*, MA, Mar. 2002.
[76]. Rita H. Wouhaybi, and Andrew T. Campbell，Phenix: Supporting Resilient Low-Diameter Peer-to-Peer Topologies，*infocom'04*
[77].Liben-Nowell, H. Balakrishnan, and D. Karger, "Analysis of the Evolution of Peer-to-Peer Networks,"*ACM PODC*,2002.
[78]. K.P. Gummadi, R. Gummadi, S.D. Gribble, S. Ratnasamy, S. Shenker, and I. Stoica, "The Impact of DHT Routing Geometry on Resiliency and Proximity," *ACM SIGCOMM*, August 2003.
[79]. Dmitri Loguinov et al, Graph-Theoretic Analysis of Structured Peer-to-Peer Systems: Routing Distances and Fault resiliency., *signalcom'03*.
[80]. Yunfei.Zhang et al, Recent Advances in Research on P2P Networks, *Tech. Report*, June 2006.