Peer-to-Peer Overlays: Issues and Trends

P. Arockia Mary
Department of Information Technology
Sudharsan Engineering College
Pudukkottai-622501

M. Radhakrishnan, Ph.D
Department of Civil Engineering
Sudharsan Engineering College
Pudukkottai-622501

ABSTRACT
Peer-to-Peer (P2P) networking is fast emerging as a disruptive business model displacing the server based networks. P2P technologies are poised to become general-purpose vehicles for building a range of applications for social networking, information delivery and personal communications in addition to the basic file-sharing services. At this evolutionary stage many important questions and problems crop up clamoring for answers and solutions with regard to techniques, scaling, compatibility and application trends. This paper intends to trace the answers to such questions in addition to providing a brief survey of the P2P practices and research directions.

General Terms
Ad-hoc Networking, Distributed Computing

Keywords
P2P, File Sharing, Scaling, Free Riding, Overlays

1. INTRODUCTION
P2P paradigm is a next generation model for distributed network services and applications. P2P applications have been deployed in various areas such as distributed computing, file sharing, storage services, conference streaming, content distribution and so on. The most popular application of P2P is file sharing. Many different file sharing systems such as Gnutella, BitTorrent, eDonkey, Kazaa, Emule are used to connect the millions of users.

P2P is not a client server model. There is no central server. In P2P system every peer acts as client (request the service) as well as server (provides the service)[1].

A P2P network can be structured or unstructured[1]. In structured P2P, the content is stored at specific location which makes the queries more efficient. It uses the DHT (Distributed Hash Table)[1][2][11] to achieve the goal. Unstructured network uses the flooding mechanism or random walk technique to send queries. Each message has a Time-To-Live (TTL)[3] value that limits its lifetime. This technique is not suitable for searching rare data. It is efficient for searching replicated data.

P2P runs in an overlay network. The overlay is a logical layer for message delivery between peers. Overlay routing mechanism is used to send message to others indirectly via other peers.

Overlay networks are constructed on top of another network. Any overlay network is constructed by the Internet peers in the application layer on top of the IP network. For end-to-end communication the overlay networks were developed. It is not dependent on Network Address Translation. In overlay network, the network service is not available within the network. An important goal in P2P network is that all clients provide resources (bandwidth, storage and computing power).

P2P overlays are self-organizing. P2P search is different from web search. Web search engine stores the index of websites in a set of servers that is queried by users. In P2P system, the index itself is distributed across the entire P2P overlay. A query must be routed to the correct peer.

The full time system administrator is not needed in P2P network. Every user is the administrator. User can control its shared resources.

Applications of P2P networks are file sharing, distributed computing, storage services, collaborative P2P applications and multimedia streaming services. File Sharing is the P2P service that the peer searches and downloads files from other peers connected to the network (Examples: Napster, Gnutella or BitTorrent). Another service provided by P2P is Distributed Computing where in the peers are aggregated to provide computing power to solve a large and computationally intensive problem. Examples are Avaki and Entropia. P2P storage service provides virtual stable storage via redundancy and aim to allow peers to continuously access files while preserving author anonymity. FreeNet is an example of such systems. Collaborative P2P applications enable application-level collaboration among peers such as instant messaging, chats and online games. Examples are NetMeeting, Groove and Jabber. Finally there is P2P multimedia streaming service which allows peers stream and multicast or broadcast audio and video to each other. Examples are PPLive, Freecast, VoP2P and P2PTV [4].

A common user may not be aware of a suitable scheme for his application among various schemes available. Each P2P scheme has its own advantages and disadvantages. In this paper we will compare the features of Napster, Gnutella, Freenet, CAN and Chord. The criteria for comparisons[3] are:

- Decentralization
- System Parameters
- Routing Performance
- Routing State
- Security
- Reliability
- Join/Leave Behavior

This paper is organized as follows. Section 2 describes the classifications of P2P networks. Section 3 discusses the overview of some of the P2P network schemes and their comparisons. Then the issues of P2P networks are discussed

in the section 4 and finally the trends are discussed in section 5.

2. CLASSIFICATIONS

In general, P2P networks are classified (1) Based on network features (2) Based on the services provided.

In the type1 the network features considered are the degree of centralization and the degree of structure. A P2P network may be centralized, decentralized but structured (Hybrid) or decentralized and unstructured (Pure). In centralized P2P networks there is a central directory which is regularly updated. This directory is used by the peers to locate the resources (Eg: Napster). In Hybrid P2P networks the topology is closely controlled and the file indices are systematically placed into peers using an algorithm (Examples: Chord, Content Addressable Network (CAN), Tapestry and Pastry). A Peer-to-Peer network in which if any single arbitrary chosen terminal entity can be removed from the network without having the network suffering any loss of network service it is called a Pure P2P network (Examples: Gnutella or KaZaa).

Type2 P2P networks are based on the services like File Sharing, Distributed Computing, Storage services, Collaborative P2P application, P2P platforms and multimedia streaming.

3. OVERVIEW OF P2P NETWORK SCHEMES

3.1 Napster

Napster was the most popular P2P application for music file sharing. This technique used a centralized server for storing an index of a file[4]. If a peer needs a particular file then it searches from a centralized server. A centralized server provides an address of the peer that holds the corresponding search file. File transfer is performed directly between two peers[5]. A centralized server provides an address only.

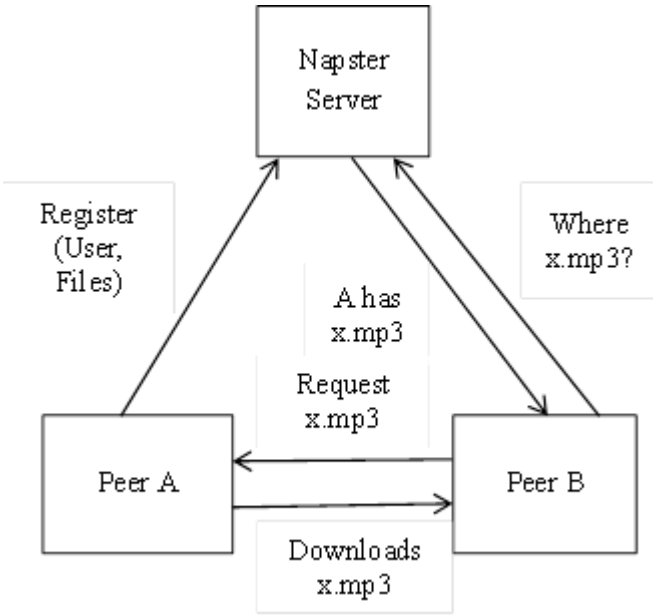

Figure 1: Napster Model

3.2 Gnutella

Peers are referred as servent[6]. A servent is a node that has both server and client functionality. The servent uses the flooding technique[3][4] for searching the files. Each message has a TTL value that limits its lifetime. The TTL value of a message is set by the source and decreased by 1 at each peer that received the message. If the TTL value of a message reaches 0 then it is no longer forwarded. It uses the following message types[3][4].

Ping – Find other host

Pong – Reply to a ping

Query – Search for a file

Query Hit – Response to a query

Push – Download a request

Bye - Tell the remote host that the connection is being closed

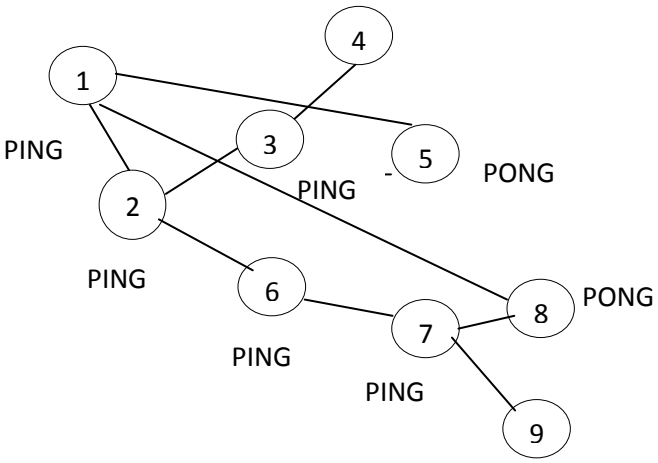

Figure 2: Gnutella Model

3.3 FreeNet

Every peer has identifiers. Peer identifiers are called as routing key. Identifiers are created using SHA-1 hashing technique[3]. Each peer has a routing table that stores links to other peers. Each entry contains the routing key of the peer. The freenet architecture is decentralized and distributed. Searching keys are passed through nodes in which each node decides the next location to send. Each request has Hop-To-Live(HTL)[3] limit to prevent infinite loops.

The node can prevent loops by rejecting requests they have seen before. This process continues until the data is found or request exceeds its Hop-To-Live limit. The result is passed to the sending node[4][7].

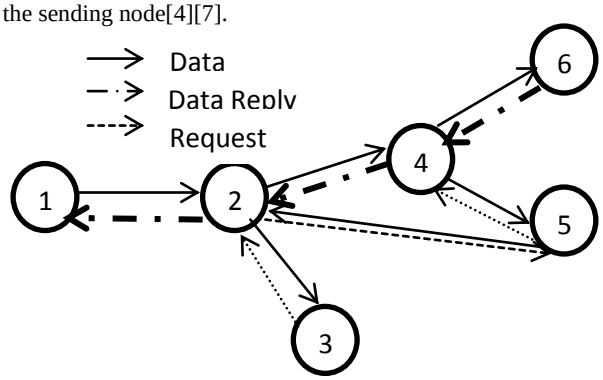

Figure 3: Example of FreeNet

Fig 3 shows a node '1' that sends the request to the node '2'. Node '2' selects the node '3' and forwards the message. Node '3' does not have a node to forward to and sends the 'request failed' message to '2'. Now '2' forwards the message to '4' and node '4' forwards to '5'. Node '5' forwards to '2' and finds a loop and returns the 'request failed' to '5'. Node '5' cannot send to other nodes and simply backtracks to '4'. Node '4' forwards the request to the next choice of the node '6'. Now, the node '6' contains the data. The data is sent back to 4, 2 and 1.

3.4 Content Addressable Network (CAN)

In centralized networks request from peers are handled by a central server which has the IP addresses of all the peers and the indices of the available files. Though the centralize model can be improved by flooding mechanism it is still unscalable. A Content Addressable Network (CAN)[2][3][8] introduces a new mechanism to improve the scalability.

CAN is a hybrid network. In which the peers are virtually located in a d-dimensional Cartesian coordinate space. The virtual space is partitioned among all the nodes giving each peer a virtual address space. Each peer maintains a part of the hash table and also information about its neighbors.

When a new peer joins the network its location address is obtained from the hash table. If a node is already present in that place it is split in half to give space for the new peer. When a request for service is received, a neighboring peer that is closest to the peer where the service is available is selected. Requests are routed through the intermediate peer to the target. CAN is scalable and can be used to implement large scale storage management and retrieval system[4].

Example: 2D space with dimension [0,80]x[0,80] is shown in figure 4. In this figure five peers are shown: A, B, C, D, X and E. Objects are stored in the space and stored at the peer whose zone assignment contains that key. For example object s with key [65, 15] would be in peer E's zone.

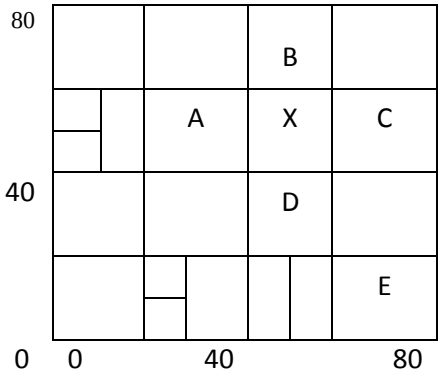

Figure 4: Peer X's coordinate neighbor set= {A,B,C,D}

3.5 Chord

Chord[2][3][8][9][10] is a structure P2P network using a ring topology[4]. It organizes peers on a logical ring and peers maintain neighbor pointers spaces at logarithmic intervals around the ring. Each peer has a link to its successor and predecessor peers on the ring. The chord routing table is called as finger table[3]. Figure 5 shows a chord ring with eight peers[4]. Peer1's finger table consists of three intervals: from peer 2 to 3, from peer 3 to 5 and from peer 5 wrapping around back to peer 1. In addition to the finger table each peer maintains links to its successor in the address space.

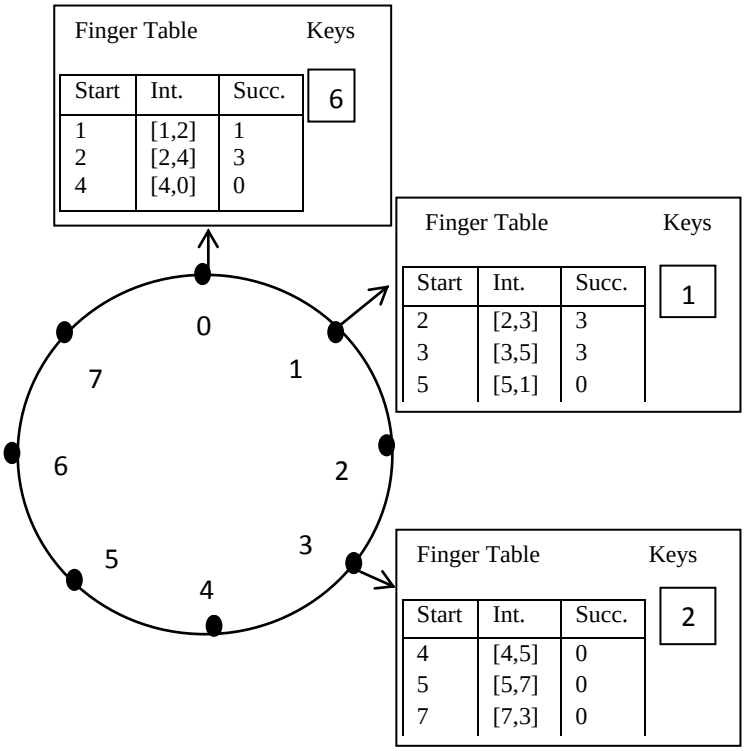

Figure 5: Chord Model

3.6 Comparisons of Napster, Gnutella, FreeNet, CAN and Chord

Table 1. A comparison of Napster, Gnutella, FreeNet, CAN and Chord[4]

	Napster	**Gnutella**	**FreeNet**	**CAN**	**Chord**
Decentralization	Centralized	Distributed	Decentralized and distributed	DHT based	
Operation Architecture	Objects are identified from central server	Queries are sent to peers through Flooding/Random Walk	Objects are identified by keywords	Multi-dimensional	Uni directional and circular space
Routing Performance	-	TTL	HTL	$O(d.N^{1/d})$	O(log N)
Routing State	Constant	Constant	Constant	2d	Log N
Security	Low	Low	Low		
Reliability	Central point of failure	No central point of failure	Performance degrades when number of peers increases	Suffers from man-in-middle and Trojan	
Join/Leave Behavior	Constant	Constant	constant	2d	$(\log N)^2$

4. ISSUES

4.1. ISP Issues

1. Firewalls, though a very important security tool, they present a serious problem to P2P network. A typical firewall prevents random nodes on the internet to connect with nodes in an internal network. A node secured like this by a firewall cannot function as a server. Thus the object of P2P overlay is defeated.

2. Dynamic IP address assignment is on the increase due to increasing modem users and its adoption by broadband providers to ensure uninterrupted service to their customers. Since a computer's IP address keeps changing every day P2P applications like file sharing and instant messaging find it tough to keep track of the address of the peer dynamically.

3. The above two issues get compounded with the use of NAT (Network Address Translator) which practically makes a destination peer node inside a network unreachable to an outside peer. One approach to circumvent this issue is to use an intermediate server in a technique called NAT Traversal. However, in a large overlay the load on the server grows and a single server may not scale adequately.

The above ISP issues are borne out of the three techniques namely, Firewall, Dynamic IP and NAT, which are meant to provide secure and scalable Internet architecture to the millions of netizens. But an undesirable fall-out has been the relegation of most computers to just clients[4][13].

But, P2P overlays demand every connected node to act as a server of resources too. The fast-expanding use of P2P applications worldwide presents a compelling need to find an effective solution to these issues.

4.2. Systematic Issues

1. Ensuring service quality is very difficult as the server resources are user-controlled. A client peer may not get the best of what is sought though available on the overlay.

2. Free-riding is a recurrent endemic issue that has no easy solution A free rider sits as a client most of the time and uses the system resources on the overlay rarely taking a server's role[14][15].

3. The overlay's ability to provide reliable and secure transactions is very low and needs to be improved to extend its use to more serious applications beyond P2PTV, VoP2P, etc.

4. An important incentive to make popular P2P applications more robust would be the availability of a well-proven revenue model that could be integrated with the application. Subscription models are gaining acceptance while other models like ad-based revenues are still under testing phase. Highly distributed nature of the P2P architecture makes designing and implementing such models very difficult.

5. TRENDS

1. For new entry application developers P2P technology is a handy testing ground due to its low barrier of entry. Once the new application is proved as a success it can be moved on as a more reliable client/server application. Examples are: Music album by young artists, Novels by new authors.

2. File sharing and instant messaging are currently the popular applications in P2P overlays. They are mostly a single application on a specific overlay. However, multiple applications on one overlay are possible. A P2P overlay can be modeled as a service delivery platform in which every peer

can offer specialized services to other peers. Services can be searched on the overlay using the network's search capability.

3. While the P2PTV is already a highly popular application its use will extend to the delivery of HDTV as well.

4. Mobile networking and Internet of Things (IoT) are two fast emerging technologies that will cause a paradigm shift in network applications soon. P2P technology as it matures will become an inevitable tool to enable such a scenario.

5. Another important emerging trend is sensor grids that find applications in air pollution monitoring, weather monitoring, healthcare monitoring and surveillance. "A new paradigm is required (for such applications) where much of the analysis is performed within the network itself using pervasive computing technologies such as the Peer-to-Peer model"[16].

6. CONCLUSION

P2P network is an overlay network for distributed object store, search and sharing. Its use by netizens is widespread and increasing. Several protocols are freely available in the open domain and their user-base is widening. Since the P2P overlay is mapped over the physical ISP network and the underlying protocols of the ISPs were not designed with P2P model in mind, there are many issues in its evolution as a standard model. Research efforts to resolve these issues are being undertaken widely and the trends, both in technology and applications, indicate that P2P might soon become a de facto network for most net users.

7. REFERENCES

[1] Eng Keong Lua, Jon Crowcroft, And Marcelo Pias, Ravi Sharma, Nanyang, Steven Lim, A Survey and Comparison of Peer-To-Peer Overlay Network Schemes, IEEE Communications Surveys & Tutorials • Second Quarter 2005

[2] Timo Tanner "Distributed Hash Tables in P2P System – A literature survey", 2005, HUT T-1 10.551.

[3] Bui Minh Nhat, "Searching in P2P Networks: A survey", TKK T-110.5190 Seminar on Networking, 2009.

[4] John F. Buford, Heather Yu, Eng Keong Lua "P2P Networking and Applications", 2009, Morgan Kaufmann Publishers.

[5]Karl Aberer, Manfred Hauswirth, "Peer-to-peer information systems: concepts and models, state-of-the-art, and future systems", ICDE 2002.

[6] Rudiger Schollmeier, "A Definition of Peer-to-Peer Networking for the Classification of Peer-to-Peer Architectures and Applications", Proceedings of the First International Conference on Peer-to-Peer Computing (P2P.01), 2002 IEEE

[7] Freenet. http://freenetproject.org.

[8] Kostas Stefanidis, A survey of peer-to-peer system, [Internet] http://courses.cs.vt.edu/cs5204/fall09-kafura/Papers/FileSystems/P2P-Survey-Kostas.pdf

[9] Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, Hari Balakrishnan, "Chord: A Scalable Peertopeer Lookup Service for Internet Applications", In Proceedings of SIGCOMM'2001, Aug. 2001.

[10] Li, Stribling, Gil, Morris, Kaashoek. "Comparing the performance of distributed hash tables under churn", February 2004

[11] K. Gummadi, R. Gummadi, Gribble, Ratnasamy,Shenker, Stoica. The Impact of DHT Routing Geometryon Resilience and Proximity. *In Proceedings of ACM SIGCOMM, Aug. 2003.*

[12] T. Klingberg, R. Manfredi. Gnutella 0.6. Technical report, June 2002. [nternet] http://rfc-gnutella. sourceforge.net/src/rfc-0_6-draft.html.

[13] Jessie Hui Wang. Chungang Wang. Jiahai Yang . Changqing An "A study on key strategies in P2P file sharing systems and ISPs' P2P traffic management", 2011, Springer.

[14] Hiroshi Nishida, Thinh Nguyen, "A Global Contribution Approach to Maintain Fairness in P2P Networks", IEEE Transactions on Parallel and Distributed Systems, Vol. 21, No. 6, JUNE 2010.

[15] Murat Karakaya, Ibrahim Korpeoglu, and Ozgur Ulusoy "Free Riding in Peer-to-Peer Networks" 2009, IEEE Computer Society.

[16] Ma. Y, Richards M, Ghanem M,Guo Y, Hassard J, "Air Pollution Monitoring and Mining Based on Sensor Grid in London", sensors(6):3601,(2008), ISSN 1424-8220; http://www.mdpi.com/1424-8220/8/6/3601