International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 04 | April - 2022 Impact Factor: 7.185 ISSN: 2582-3930

A Survey of Applications for Peer-to-Peer File Transfer

Khushi Kulkarni[1], Aditi Mishra[2], Shantanu Kshirsagar[3], Mohd. Amin Shaikh[4], Nilesh Kulal[5]

[1, 2, 3,4] BE Computer Engineering Student, Terna Engineering College, Nerul, Maharashtra, India

[5] Professor, Dept. of Computer Engineering, Terna Engineering College, Nerul, Maharashtra, India

---

ABSTRACT

This survey paper explores various methodologies and available P2P file sharing platforms. Such architectures and systems are characterized by direct access between peer computers, rather than through a centralized server. The recently formed Peer-to-Peer Working Group, a consortium including industry leaders aiming at the advancement of infrastructures and best-known practices for peer-to-peer computing, defines p2p as the "sharing of computer resources by direct exchange". Apart from resources, p2p offers a way of decentralizing administration (as well as cost). A peer-to-peer network allows computer hardware and software to communicate without the need for a centralized server. P2P file sharing allows users to access media files such as books, music, movies, and games using a P2P software program that searches for other connected computers on a P2P network to locate the desired content.

Key words: P2P network, File sharing, Simplicity, Desktop app, Convenient.

1. INTRODUCTION:

Peer-to-Peer File Sharing systems are no longer just a new fad technology. They have become ingrained in our Internet culture.
Peer-to-peer (P2P) is a decentralized communications model in which each party has the same capabilities and either party can initiate a communication session. Unlike the client/server model, in which the client makes a service request and the server fulfils the request, the P2P network model allows each node to function as both a client and server. Peers are equally privileged, equipotent participants in the application.
The individual users in this network are referred to as peers.
File sharing is the dominant p2p application on the Internet, allowing users to easily contribute, search and obtain content. Grid computing, which has emerged as a field distinguished from conventional computing by its focus on wide area distributed computing, large-scale resource sharing and problem solving is closely related to p2p. It is expected that there will be an even stronger convergence between them as p2p technologies become more sophisticated [1].

Online File transferring apps are required because:

- You can manage your documents and send your files anywhere and anytime.
- Distributing documents is not so complicated with a file sharing software.
- Simplified administration as it keeps files maintained and organised.
- Greater security compared to physical file transfers.
- Saves time and eliminates multiple file versions.

1.1 Drawbacks of other file sharing platforms:

- **FTP:** Servers can be spoofed to send data to a random port on an unintended computer. Also, Inconsistent due to inability to track what has been uploaded on the remote system,
- **Server-client:** If all the clients simultaneously request data from the server, it may get overloaded. This may lead to congestion in the network. If the server fails for any reason, then none of the requests of the clients can be fulfilled.
- **Cloud based:** Cloud based storage is dependent on having an internet connection. If you are on a slow network, you may have issues accessing your storage. Also, there are additional costs for uploading and downloading files from the cloud.

1.2 Reasons of choosing a P2P file sharing network

1. Using a secure peer to peer connection and its data channel huge files can be transferred without storing it on any server. Making it really robust and truly private as only the connected clients/peers are communicating directly with no middle server for transfers making it really robust and truly private as only the connected clients/peers are communicating directly with no middle server for transfers.
2. P2P Systems are more resilient to single-point-of failure than client-server model as there is no central server. Failure of a single peer does not affect the network.
3. The responsibilities of the central server are distributed among each individual peer of the network. Every user is an administrator of his machine and have control on his shared resources.
4. P2P network is a logical overlay for the underlying physical systems, so the communication in overlay network is through virtual channels which avoid the

obstacles from firewalls and Network Address Translation (NAT).

1.3 Characteristics of P2P network:

- *Nature of node*: The nodes are symmetric in nature that means every node in the network can act as client (raising requests) as well as server (serving queries).
- *Scalability:* There is no restriction on number of participating peers where as in traditional networks the number of participating nodes depends on the capacity of server [7].
- *Heterogeneity:* The participating machines are not necessarily homogenous. A P2P network may have a very slow machine and a high end super computer working together.
- *Attacks:* The heterogeneous peers make virus and worms harder in the network. Thus, P2P is resilient to attacks.
- *Dynamism:* In P2P applications we find dynamic change of topology due to joining of new nodes or leaving of existing nodes from the network.
- *Self Organization*: The nodes of the network reconfigure according to the dynamic changes in the topology due to a node joining or leaving [7].
- *Fairness:* Each participating machine should contribute resources to the network based on its capacity [8].
- *Huge Resources:* In P2P, we will have large collection of resources due to voluntary participation of millions of simultaneous users from all over the world.
- *Flexibility:* As there is no central controlling system, each participating peer is completely flexible making the overall system unreliable.
- *Performance:* To avoid single-point-of-failure, data and object references are replicated at distinct peers. This also balances access load and enhances search & retrieval of data.

2. CLASSIFICATION OF P2P FILE SHARING APPLICATIONS

In this Section we examine two general aspects of p2p architectures, according to which the p2p systems can be differentiated and categorized: The degree of centralization, and the network structure.

2.1 Degree of centralization

P2p file sharing architectures can be classified by their "degree of centralization", i.e. to what extent they rely to one or more servers to facilitate the interaction between peers. Three categories are identified:

- **Centralized P2P architecture:** This is the first generation of P2P, where there will be one or more central servers. Unlike traditional client-server model, here the servers contain only the Meta information about the shared resources (e.g.: only node ID or address where the shared content is available) instead of maintaining/storing the actual resources. The peer in the network has to initially raise query request for resources to the central server, the server which is having the meta information replies with the list of peer ID"s who can provide services to the request.
- **Purely decentralized p2p architectures**: All nodes in the network perform exactly the same tasks, acting both as servers and clients, and there is no central coordination of their activities. The nodes of such networks are termed "servents" (SERVers+clieENTS).
- **Partially centralized systems:** The basis is the same as with purely decentralized systems. However, some of the nodes assume a more "important" role than the rest of the nodes, acting as local central indexes for files shared by local peers. These nodes are called "Supernodes", and the way in which they are selected for these special tasks vary from system to system. It is important to note that these Supernodes do not constitute single points of failure for a p2p network, since they are dynamically assigned and in case they are subject to failure or malicious attack the network will take action to replace them with others.
- **Hybrid P2P:** There is a central server facilitating the interaction between peers by maintaining directories of the shared files stored on the respective PCs of registered users to the network, in the form of meta-data. The end-to-end interaction is between two peer clients. However, these central servers facilitate this interaction by performing the lookups and identifying the nodes of the network where the files are located. The terms "peer-to-peer" or "broker mediated" are sometimes used for such systems [2].

2.2 Network Structure

P2p systems constitute highly dynamic networks of peers with complex topology. This topology creates an overlay network, which may be totally unrelated to the physical network that connects the different nodes (computers). P2p systems can be differentiated by the degree to which these overlay networks contain some structure or are created ad-hoc. By structure here we refer to the way in which the content of the network is located with respect to the network topology

- **Unstructured networks:** In unstructured networks the placement of data (files) is completely unrelated to the overlay topology. Since there is no

information about which nodes are likely to have the relevant files, searching essentially amounts to random search, in which various nodes are probed and asked if they have any files matching the query.

- **Structured networks:** They have emerged mainly in an attempt to address the scalability issues that unstructured systems are faced with. The random search methods adopted by unstructured systems seem to be inherently unscalable [3], and structured systems were proposed, in which the overlay network topology is tightly controlled and files (or pointers to them) are placed at precisely specified locations. These systems provide a mapping between the file identifier and location, in the form of a distributed routing table, so that queries can be efficiently routed to the node with the desired file [3]. Structured systems offer a scalable solution for exact-match queries, i.e. queries in which the complete identifier of the requested data object is known (as compared to keyword queries). There are ways to use exact exact-match queries as a substrate for keyword queries [4], however it is not clear how scalable these techniques will be in a distributed environment.
- **Loosely structured networks:** are in between the two. File locations are affected by routing hints, but they are not completely specified, so not all searches succeed [3].

2.3 Classification of commonly used P2P file sharing application:

	Unstructured Networks	Loosely structured Networks	Structured Networks
Hybrid Decentralized	Napster		
Pure Decentralized	Gnutella	Freenet	Can, Chord
Partially Decentralized	Kazaa		

3. OVERVIEW OF A BASIC SYSTEM

This section discusses the basic and common architecture of a file sharing application using the concepts of P2P.

3.1 Architecture Diagram

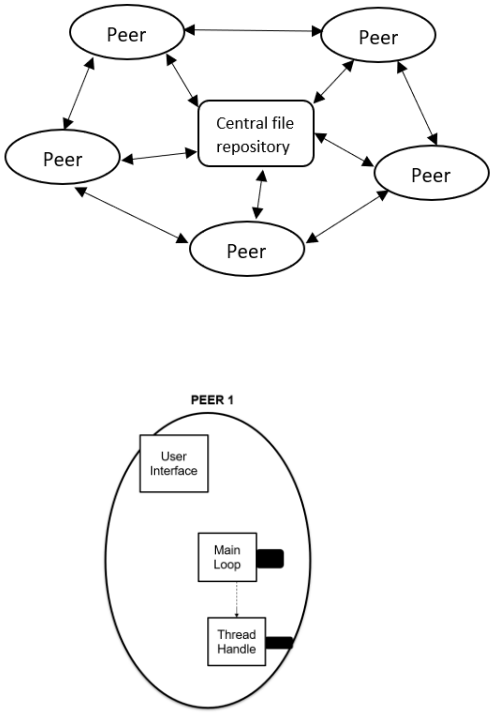

Figure 1. Architecture diagram of a centralized P2P system

The Peer module: manages the overall operations of a single node in the P2P network. It contains a main loop that listens for incoming connections and creates separate threads to handle them. The Peer module for various message types and the main loop would dispatch incoming requests to the appropriate handler. The Peer is initialized by providing a port to listen for incoming connections and node identifier.

The PeerConnection module: encapsulates a socket connected to a peer node. The framework currently uses TCP/IP sockets for communication between nodes. A PeerConnection object provides methods that make it easy for the peer to send and receive files and acknowledgments in the P2P algorithm.

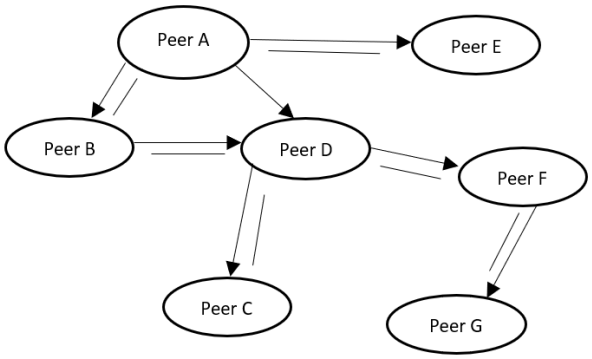

Figure 2. Architecture diagram of decentralised (unstructured) P2P network

Figure 2 shows Unstructured P2P network using flooding-based routing strategy. In this network new node joins by forwarding ping message to the existing nodes which responds through pong message [6]. The joining node establishes some among the replied nodes as its neighbours. Figure 2 shows a sample network how searching takes place through flooding. Suppose Peer A requires data which is available with Peer D, therefore Peer A starts broadcasting its request to its neighbours and gradually to others peers in the whole network.

In Unstructured P2P [5], it is difficult to predict which peer is having data (in case of search requests) and there will be no guarantee for completeness of answers (as peers having the data may or may not be available in the network at the particular moment). The response time may also vary for the same request at distinct times.

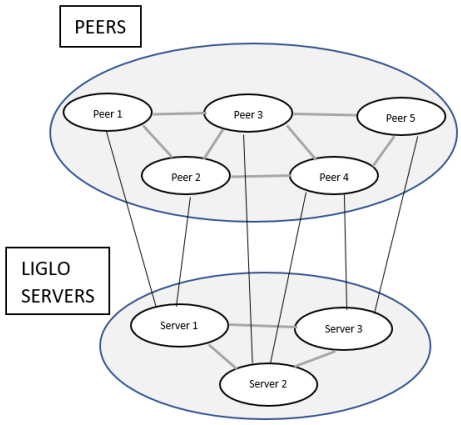

Figure 2. Architecture diagram of Hybrid P2P network

In the above example of hybrid P2P system, just like super peers it maintains Location independent global names lookup servers (LIGLOs) in the upper layer. The underlying peers communicate with the corresponding LIGLOs for the queries. A new node can join to any number of LIGLOs and chose its neighbours [9].

3.2 Information Architecture Diagram

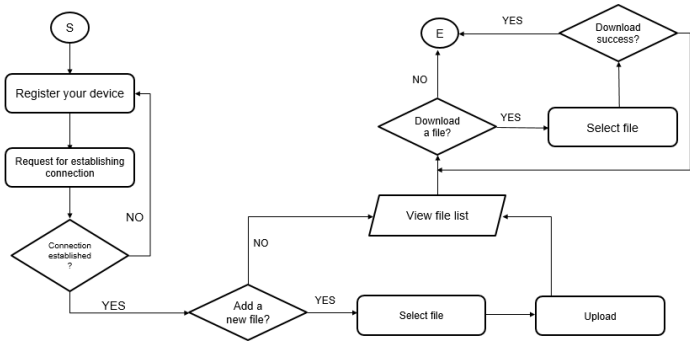

Figure 2. Information Architecture

A peer can request or send a file to another peer in the network. Also, it can view the files already existing files and download.

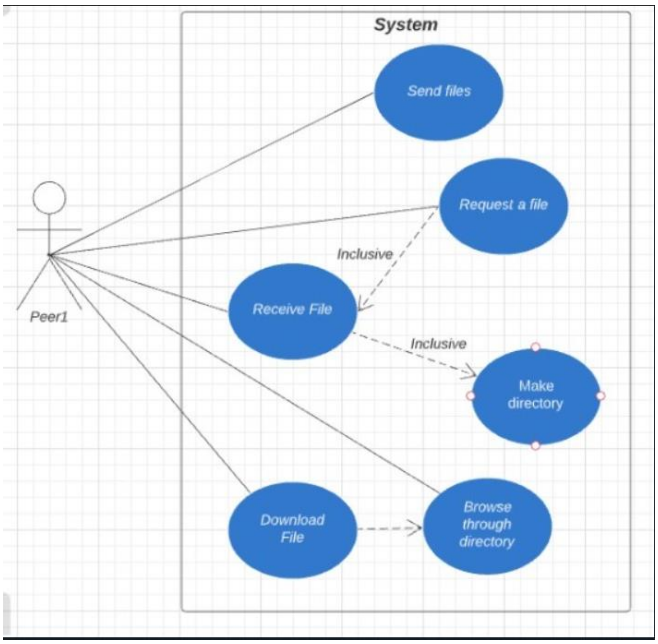

Figure 3. User case diagram

4. DESIGN ISSUES AND CHALLENGES OF P2P FILE NETWORK

- *Resource discovery*: Locating the resources as per the user requirements is becoming more and more important with the emergence of huge amounts of information and growth size of P2P networks. Especially in Unstructured P2P networks, due to the feature of dynamic and loosely coupling relations among peer nodes, the data resources resided in peers are usually independent of logical topology and thus making the resource discovery strategies hard to work. There are many methods for finding the resources in unstructured P2P, some of them are flooding, Random Walk algorithm, Gossip based Search mechanism [10], and Query Routing Trees (QRT) [11] etc.
- *Availability of data/resources:* A node can join or leave the P2P network at any instant of time based on its interests. As the nodes are autonomous it is difficult to predict whether the required data/ resources are available at a particular moment. At some situations the data may available but not complete. This is because in P2P the data is divided and shared among distinct peers whose availability is not guaranteed. One of the basic solutions for such

problem is Replication i.e. Duplication of data at multiple peers.

- *Data consistency*: Continuously relying on Replication for availability of data and for reducing query processing time not only increases storage cost but also creates data consistency problem. If any change is done in the data, then it has to be reflected in all its, without which the outdated copies of data may be served to the user requests. So, it is the basic task of the network to maintain consistency among all its replicas. It can be done by removing outdated copies or by applying few methods to update replicas immediately when change occurs at original/owner node (it is a costlier process). As an alternative approach, the nodes having replicas should check with the node having original data at regular time intervals for any updations. If any updations are there, then it will be indicated with version numbers and thus consistency has to be maintained.
- *Load Balancing:* It means distribution of workload on multiple peers in the network to achieve optimal resource utilization for maximizing throughput and minimizing overall response time. If in the network nodes are homogeneous then the workload is equally distributed, but P2P allows heterogeneous nodes so in that case the distribution of load should be based on the capacity of the particular node participating in the network. As nodes join and leave the network at high rate the load has to be balanced dynamically. Thus, load balancing in P2P forms a major challenging issue such that no node should be overloaded or underloaded with work. Under Static methods, the load has to be transferred from heavily loaded node to lightly loaded node and vice versa at the times of node joining and node leaving respectively.
- *Congestion:* Congestion of traffic is mainly seen in P2P file sharing applications because of which packet (data is transferred as packets) loss may occur or may create a delay in process. Some of the factors that create congestion (Overflow of traffic) are attacks and through relying on same route for several requests. In order to control congestion in P2P, we should monitor on attackers who increases traffic by flooding and at the same time we should find optimal routes for the transfer of packets.
- *Security:* Security is more complex in P2P systems than in client server model, because of openness and autonomous nature of P2P. Unlike traditional client-server model, in P2P internal data is exposed to its fellow nodes so easy for the attackers to hack the data. A malicious node can easily enter the network and can create Denial of Service (DoS) [20][21]. The Commonly seen security issues [12] on Peer – to - Peer networks are Routing attacks, Retrieval attacks, Distributed DoS (DDoS) attacks, privacy & identity, poisoning the network etc. Routing Attacks are mainly seen in Structured P2P where routing is through routing tables. Attackers will wrongly updates the tables or forward the query not as in routing table (incorrectly). DoS attacks can be seen at network layer by flooding traffic and at application layer by creating huge application requests.

5. CONCLUSION

With the rapid increase of P2P enabled file sharing applications over the internet there is an increased focus on P2P Networks by the researchers. In this paper, we have drawn an overview about the various P2P system architectures and the challenging issues that have to be taken care while working with P2P systems. The paper also briefly compares P2P method with other file sharing networks.

REFERENCES

[1] Stephanos Androutsellis-theotokis. A Survey of Peer-to-Peer File Sharing Technologies (2002).
[2] HC. Kim. P2p overview. Technical report, Korea Advanced Institute of Technology, August 2001.
[3] Q.Lv, S.Ratnasamy, and S.Shenker. Can heterogeneity make gnutella scalable? In Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02), MIT Faculty Club, Cambridge, MA, USA, March 2002.
[4] IH. Witten, A.Moffat, and TC. Bell. Managing Gigabytes: Compressing and Indexing Documents and Images. Morgan Kauffman, second edition, 1999.
[5] XianfuMeng&Dongxu Liu, "A traffic-efficient message forwarding approach in unstructured P2P networks", Springer publication, DOI 10.1007/s12083-015-0326-2.
[6] D Giridhar Kumar, Ch D V SubbaRao, "Peer – To – Peer Computing: Architectures, Applications and Challenges",International Journal of Recent Technology and Engineering (IJRTE).
[7] B Lalitha, Ch D V SubbaRao, "Peer-to-Peer Systems: Taxonomy and Characteristics", International Journal of Computer Science and Technology.
[8] AlirezaNaghizadeh, BehroozRazeghi, EhsanMeamari, MajidHatamian, Reza EbrahimiAtani, " C-trust: A trust management system to improve fairness on circular P2P networks", Springer publication, DOI 10.1007/s12083-015-0402-7.
[9] Wee Siong Ng, Beng Chin Ooi, Kian-Lee Tan. "BestPeer: a self-configurable peer-to-peer system."s, IEEE 2002.
[10] HE Ming, ZHANG Yujie and MENG Xiangwu, "GossipBased Resource Location Strategy in Interest Community for P2P Networks", Chineese Journal of Electronics, Vol.24, No.2, April-2015.
[11] Gong Weihua, Jin rong, YANG Lianghuhai and HUANG Decai, "Enhanced Entropy-BasedResource Searching in Unstructure P2P Networks", Chineese Journal of Electronics, Vol.24, No.2, April 2015.
[12] Security issues: https://www.cse.wustl.edu/~jain/cse571-07/ftp/p2p/.