

Computer Systems 2  
CEN 502  
Project 2 Survey

Group Members:

Mayank Gupta

ASU ID: 1209469518

Yogesh Kubal

ASU ID: 1209536364

Aman Karnik

ASU ID: 1209536091

## Introduction:

Peer to Peer File Networking is the partition of tasks or workloads between peers wherein each peer is equally privileged and equipotent. P2P networking is an alternative to the traditional client-server networking approach. It has several benefits such as increased bandwidth, robustness, efficiency, resource provisions by peers and so on. The utilization of resources can be of utmost importance in the sharing of large files. P2P design has evolved through many stages starting from the early models of Napster to the currently global BitTorrent protocol model.

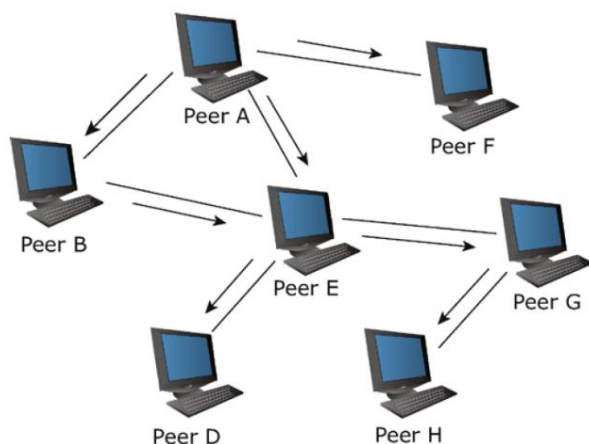
A brief survey of three such existing P2P technologies denoting the architectures and the protocols on which they operate is stated below:

### I) Fully Decentralized P2P Architecture

The clients and the servers are all equipotent and equally privileged such that there is no central authority governing the entire network. The nodes (clients and servers) connect to each other using some specific software application. BitTorrent and Gnutella are two widespread examples of protocols which employ the fully decentralized architecture.

#### BitTorrent:

BitTorrent is one of the most widely used P2P file sharing protocol. Its implementation is effective because it is safe, fast, can handle and distribute large amounts of data over the Internet. BitTorrent is one of the most common protocols used for transferring large files among peers. BitTorrent utilizes a pure peer-to-peer concept in which the peers themselves are the only entities<sup>[1]</sup>. BitTorrent bases its operation around the concept of a torrent file, a centralized tracker and an associated swarm of peers.



BitTorrent protocol emphasizes the following characteristics:

- Decentralization: the system is collectively created and maintained by the nodes without any central coordination.
- Scalability: the system performs efficiently even with thousands or millions of nodes.
- Fault tolerance: the system is reliable even with nodes continuously joining, leaving or failing.<sup>[1]</sup>

## Operating Protocol:

The operation of BitTorrent is elementary such that the users need to have a BitTorrent client installed on their machine which runs the BitTorrent protocol. There are a variety of clients available at the disposal of the user some of which are uTorrent, Deluge, Xunlei, Vuze etc.

Each client is capable of requesting and transferring any type of computer file over a network using a protocol. A peer can be defined as any computer running as a client. A peer will first create a small file called as a torrent which will hold some metadata about the files to be shared and tracker information wherein the tracker is the main computer that will coordinate the file distribution. Peers that wish to download a file will need to first download a torrent file for it and run it using a BitTorrent client, connect to a specific tracker which will branch out the list of peers from whom the chunks of data can be downloaded<sup>[2]</sup>. A peer connects with other peers to download and simultaneously upload chunks of data using different IP connections and the file demanding peer will check the list of chunks of data held by the neighboring peers and download in a manner called “rarest first” such that the rarer data blocks are downloaded from the neighboring peers so that data is redistributed evenly. A peer will also need to service requests that it receives. BitTorrent uses a clever trading algorithm wherein the peer will service neighbors that are supplying data at the highest rate. Four neighbors sending data at the highest possible rate are selected and they are tagged as unchoked neighbors. Every 30 secs an additional neighbor is connected and this neighbor is called as optimistically unchoked. The process continues and the four high rate providing neighbors will change over time. Thus at a given time five neighbors (four top and one randomly selected probing peer) will be supplied with data chunks whereas the remaining neighbors will not be provided with any data i.e. the neighbors are referenced as choked.

This is the case considering just one peer, a BitTorrent ecosystem consists of millions of peers all interconnected in a network sharing files amongst each other.<sup>[3]</sup>

The BitTorrent protocol selects pieces by using the following four simple policies:

- Policy #1: Strict Policy: Until a piece is assembled, only download sub-pieces for that piece.
- Policy #2: Rarest First: Determine the pieces that are most rare among your peers and download those first.
- Policy #3: Random First Piece: Select a random piece of the file and download it.
- Policy #4: Endgame mode: When all the sub-pieces that a peer doesn't have are actively being requested, these are requested from every peer.<sup>[1]</sup>

When a peer acquires the entire file it can continue seeding the torrent by uploading chunks of data to other peers or it may selfishly leave the torrent. Moreover, any peer can leave the torrent at any given time and return back later to resume from the previous position.

## Conclusion:

BitTorrent has rocketed to become the most popular P2P protocol for networking in just a few years primarily because of its simplicity and the enhanced capabilities. The protocol moved from implementing a hybrid peer-to-peer scheme (a centralized tracker) to a pure peer-to-peer scheme. BitTorrent eases file sharing considerably and downloads pieces of the files using algorithms that are embedded into the protocol itself. BitTorrent is an upcoming technology that change the face of networking and file sharing and with added improvements this technology is going to keep on getting better and better. The first pure P2P protocol called Gnutella also operated on the decentralized architecture. There have been many clients implementing the Gnutella infrastructure with some of them discontinued and certain others still active. Discontinued P2P file sharing software such as Limewire and BearShare operated on this protocol platform whereas Shareaza, a file sharing client still currently active implements a lot of functionality that other Gnutella clients do not.

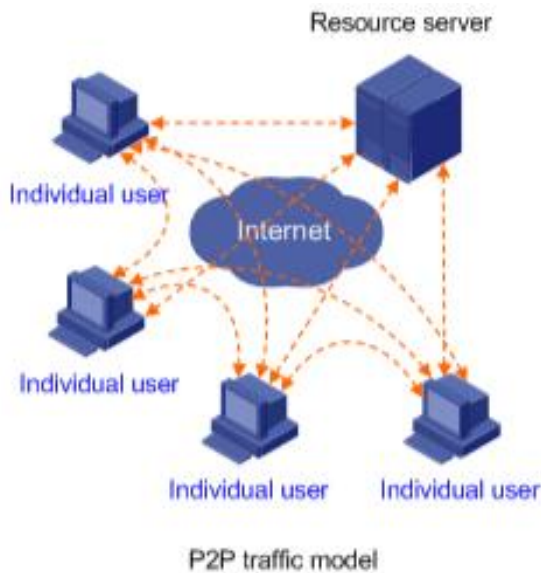
## II) Centralized P2P Architectures

A centralized P2P architecture implements a lot of functionalities of the traditional client-server architectures. There is one central server and several peers connected in a network. Direct Connect is a protocol that operates on this centralized architecture.

### Direct Connect (DC)

Direct Connect is a peer-to-peer file sharing protocol wherein clients connect to a central hub and can download files directly from other peers.

### Operating Protocol:



The central server in the network will maintain metadata of a list of files shared by peers in the network. The metadata may be a tuple which holds the ID of the object and the ID of the peers who hold the required object. Any query in the network is first directed towards the central server and the server will return a list of nodes containing the objects. The node requesting the file will initiate direct communication with the peers that hold the file. The use of the server hub is to answer queries and provide a list so that further P2P communication is established. Some hubs may also offer added functionalities such as chat capabilities and user authentication.

A centralized P2P architecture such as Direct Connect will speed up the process of resource location and it will guarantee to find all possible nodes that contain the entry. Moreover, the entire network is easier to access and navigate due to the use of a centralized server. <sup>[4]</sup>

The only issue that can hamper the entire network is that the server becomes the one single point of failure.

An important factor of a centralized architecture is improving the security and robustness of the overall network.

A classic example of a direct connect protocol implementation using the centralized P2P architecture is Napster, the music file-sharing application.

### Properties of Napster employing the Direct Connect protocol:

**Scalability:** All queries need to be routed to the central server first and all servers need to connect to the central server. Since the server has limited capabilities the server may get flooded with queries and thus a direct connect protocol system is not scalable.

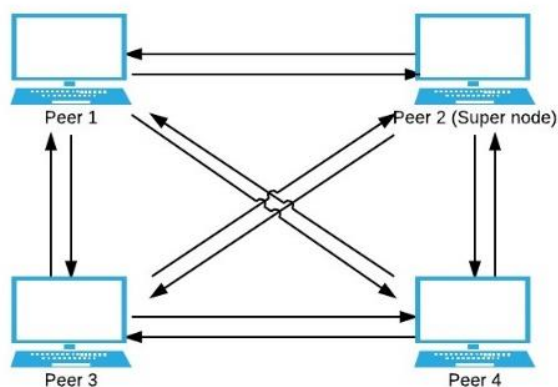
**Efficiency:** A direct connect protocol can accelerate resource locations with cheaper cost and a high efficiency.

**Security:** Since Direct Connect uses a central server which creates a single point of failure and a target for malicious attacks on the overall system. Using the server database, the IP address list of the entire peer network can be compromised. Thus in terms of privacy, security and anonymity a DC is not at all effective. <sup>[4]</sup>

### Conclusion:

Direct Connect protocol has been a set standard architecture which formed the basis for file sharing between peers with the advent of Napster. Napster on a whole didn't survive much, but there is a lot of research and continued interest to the betterment of the protocol. A successor to the hugely popular Direct Connect called Advanced Direct Connect is being developed.

## III) Hybrid Peer-to-Peer Architecture



Hybrid peer-to-peer architecture emerged as a system that integrates the advantages of both centralized and decentralized architectures. The added benefits are that the hybrid P2P architectures have some elegant mechanisms that allow them to search the resource locations faster. Hybrid P2P architectures consist of some peers which have higher processing capabilities. Such peers are addressed as 'supernodes' and they form the upper layer of the hybrid architecture.

A supernode will be in-charge of a few peers in its neighborhood. The common peers can utilize more services due to the presence of these supernodes. Moreover a hybrid system consisting of supernodes will distinguish itself from a centralized system in the following way such that the supernodes will not have as much power as a centralized server holds, and the peer will have to coordinate the operations among the peers under its supervision but also at the same time contribute its own resources and be a part of the file sharing process.

### **Operating Protocol:**

The operation of a hybrid P2P system can be stated as a peer joins into a new network, it randomly chooses a set of nodes as its neighbors and issues a query for any desired file or object or answers queries from other nodes. A peer connects to other peers in a manner such that the neighbors can be beneficial in the near future. This characteristic is called as self-configurability of a peer. As time passes queries are answered quickly, efficiently and precisely due to the dynamic nature of the configuration. Hybrid P2P systems have many advantages over standard architectures being an overall improved response time, avoiding a single point of failure and lastly optimizing the overall network. <sup>[4]</sup>

### **Conclusion:**

Hybrid P2P systems are gaining widespread popularity because of its implementation of the best features of both centralized and decentralized systems. Many newer versions of existing protocols such as Gnutella, BestPeer and so on. Even newer versions of Napster utilized some hybrid architecture. A very famous protocol implementing the hybrid architecture is FastTrack. FastTrack worked on the functionality of supernodes so as to improve the overall scalability. The developers who invented FastTrack went on to create Skype, the global leader in voice and video calls.

## References:

- World Wide Web

[1]. J.A. Pouwelse, P. Garbacki, D.H.J. Epema and H.J. Sips. A Measurement Study of the BitTorrent Peer-to-Peer File-Sharing System. Delft University of Technology.

<http://www.pds.ewi.tudelft.nl/reports/2004/PDS-2004-003/>

[2] Rüdiger 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, IEEE (2002)

[3]. Computer Networking: A Top Down Approach (6<sup>th</sup> Edition), Authors: Kurose and Ross, Chapter 2: Application Layer, 2.6 Peer to Peer Networks, Page No. 149-151

[4]. Peer-to-Peer Computing: Principles and Applications, Authors: Quang Hie Vu, Mihai Lupu, Beng Chin Ooi, Chapter 2: Architectures of Peer-to-Peer Systems, Page No. 22-24