# Communication Networks
# Project 1 - Phase 1

Shivani Chepuri 2018122004
Gowri Lekshmy 20171053

## Project topic:
Modeling and Performance Analysis of BitTorrent-Like Peer-to-Peer Networks

## Objectives:
- Studying and Understanding (the research paper) "**Modeling and Performance Analysis of BitTorrent-Like Peer-to-Peer Networks**".
- Literature Survey (on P2P networks and BitTorrents) within the scope of the main paper.

Main paper link: http://conferences.sigcomm.org/sigcomm/2004/papers/p444-qiu1.pdf
Main Paper Title: **Modeling and Performance Analysis of BitTorrent-Like Peer-to-Peer Networks**

## Literature Survey Outline and Scope:
Besides studying and understanding the above research paper, we want to look into the topics related to P2P networks that are abstracted and cut short in it. In regard to the topics and insights mentioned in the main research paper, we want to explore more papers on BitTorrent and P2P networks, keeping the scope limited to the ideas in the main research paper.

More Research papers (and References) for Survey (Study of each paper is subjected to the scope of the main paper and the project objective):

**"Fairness, incentives and performance in peer-to-peer network"**
http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=C49D7A1A99BDFB6C1C3B5D1C2640F07E?doi=10.1.1.121.9069&rep=rep1&type=pdf

**"Performance Analysis of BitTorrent and Its Impact on Real-Time Video Applications"**
https://www.researchgate.net/publication/220717176_Performance_Analysis_of_BitTorrent_and_Its_Impact_on_Real-Time_Video_Applications

**"Performance analysis of BitTorrent protocol"**
https://www.researchgate.net/publication/261269751_Performance_analysis_of_BitTorrent_protocol

**"A Performance Study of BitTorrent-like Peer-to-Peer Systems"**
https://web.njit.edu/~dingxn/papers/BT-JSAC.pdf

## Abstract of the paper:

- Develop simple models to study the performance of BitTorrent
- Present a simple fluid model and study the scalability, performance, and efficiency of such a file-sharing mechanism, Steady state network performance.
- Consider the built-in incentive mechanism of BitTorrent and study its effect on network performance.

## Approach and Workflow:

- Understand P2P networks, their architecture, and their mechanism.
- Study and understand the terminology, architecture, and working of a BitTorrent.
- Understand a simple Fluid Model and Steady-state performance
- Explore various file-sharing mechanisms and algorithms in BitTorrent or P2P networks.
- Explore Free-Riding and Optimal Unchoking in BitTorrent or P2P networks.

## Important Terms:

**Peers:** Equally privileged, equipotent participants in the application
**Client Peer:** Peer that is sending a request for a file/object
**Server Peer:** Peer that is accepting a request for a file/object, from a Client Peer
**Torrent:** A collaborative file sharing process
**Swarm:** The set of all peers that takes part in a torrent
**Seed:** A peer in a swarm that has the complete content file
**Downloader/Leech:** A peer that has only part of the file and wants to download the rest
Swarm = Seeds + Leeches
**Tracker:** A central node that tracks the operation of the swarm

## <u>Subtopics to be explored:</u>

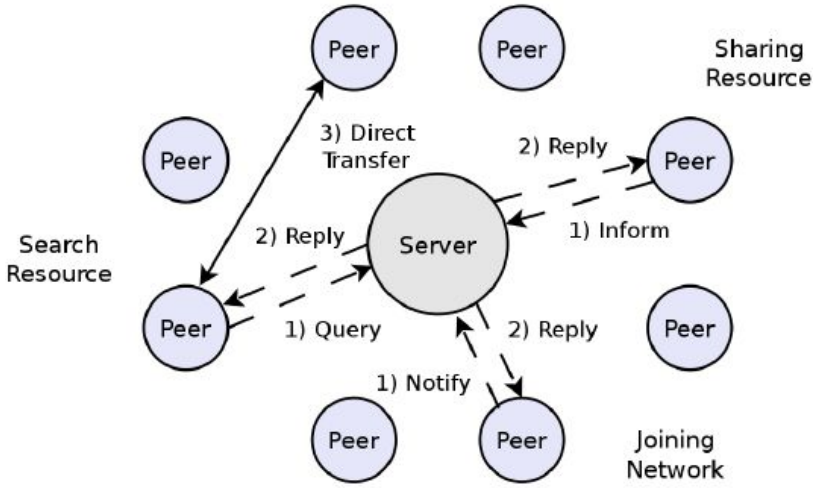
## P2P File Sharing:

- Since lists of peers may grow and shrink, the paradigm needs to keep track of loyal peers and the location of the files. This is done by dividing the P2P networks into two categories: **centralized** and **decentralized**.

## P2P Architecture:

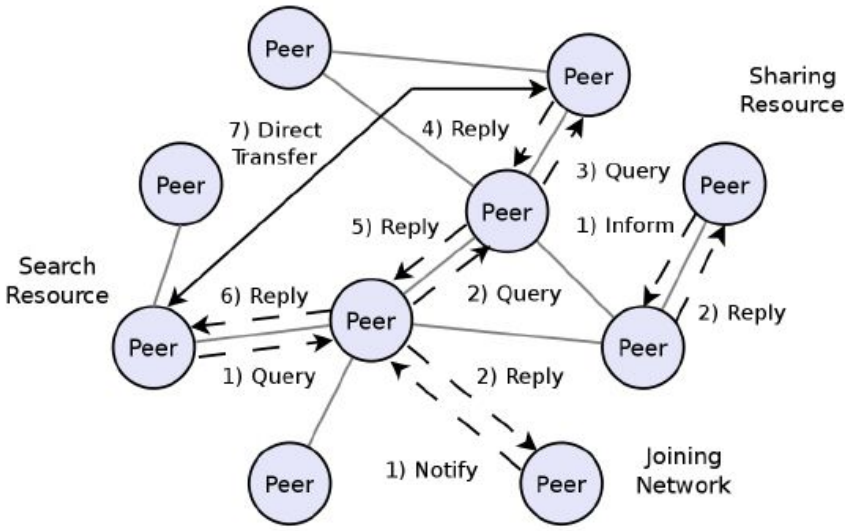
1. **Centralized P2P (Hybrid P2P) Network:**
   The directory system (listing of the peers and what they offer) used a client-server paradigm. Storing and downloading of the files are done using the peer-to-peer paradigm.
   Eg: Napster

Diagram of a Fully Centralized P2P architecture:




2. **Decentralized Network:**
   Peers are arranged into an overlay network (A logical network above the physical network).
   Depending on how the nodes in the overlay network are linked, a decentralized P2P network is classified as either **unstructured** (nodes linked randomly. Eg. Gnutella, Freenet, etc) or **structured** (predefined set of rules to link nodes, Uses Distributed Hash Table. Eg. BitTorrent).

   Diagram of a Fully Centralized P2P architecture:



## BitTorrent:

- A P2P protocol for sharing large files among a set of peers called a swarm.
- File sharing is done in a collaborating process called a **torrent**.
- A BitTorrent can be **implemented with a tracker being centralized or in a decentralized** fashion where a tracker's functions are distributed among the nodes.
- A downloader can upload data to other peers even though it may only have parts of a file (η parameter, effectiveness of file sharing)

- **BitTorrent protocol** applies a set of policies
    - To provide fairness
    - To encourage the peers to exchange pieces with other peers
    - To prevent overloading a peer with requests from other peers
    - To allow a peer to find peers that provide better service.

- The typical value of the number of concurrent neighbors is four.
- Two types of peers :
    - **Downloaders :** Peers who only have a part (or none) of the file
    - **Seeds :** Peers who have all the pieces of the file but stay in the system to allow other peers to download from them.
- The **unchoked group** is the list of peers that the current peer has concurrently connected to. It continuously uploads and downloads pieces from this group.
- The **choked group** is the list of neighbors that the peer is not currently connected to but may connect to in the future.

### Optimistic Unchoking Strategy:

To allow a newly joined peer, which does not yet have a piece to share, to also receive pieces from other peers, every 30 seconds a peer randomly promotes a single interested peer, regardless of its uploading rate, from the choked group and flags it as unchoked.

### Rarest first Strategy:

A strategy to provide a balance between the number of pieces each peer may have at each moment. A peer tries to first download the pieces with the fewest repeated copies among the neighbors so that the pieces are circulated faster.

### Endgame Mode:

- Any applied algorithm for downloading the last few pieces of a torrent.
- A protocol that is used to prevent users who have all but a few pieces from waiting too long to finish their download.
- In typical client operation, the last download pieces arrive more slowly than the others. This is because the faster and more easily accessible pieces should have already been obtained.

- In order to prevent the last pieces from becoming unobtainable, BitTorrent clients attempt to get the last missing pieces from all of their peers. Upon receiving the last pieces a cancel request command is sent to other peers.

## Free-Riding:

- Free-riding means that a peer does not contribute anything to the system, while it attempts to obtain service (or downloading) from other peers.
- If peers have global information, the free-riding problem can be solved by not uploading to peers with zero uploading bandwidth.
- In reality, peers use optimistic unchoking to explore the network and this gives an opportunity to free-ride.

## Simple Fluid Model:

Model Parameters:

**x(t)** number of downloaders (also known as leechers) in the system at time t

**y(t)** number of seeds in the system at time t

**λ** the arrival rate of new requests

Assumption: Poisson Process

**μ** the uploading bandwidth of a given peer

Assumption: All peers have the same uploading bandwidth

**c** the downloading bandwidth of a given peer

Assumption: All peers have the same downloading bandwidth, **c>=μ**

**θ** the rate at which downloaders abort the download.

Assumption: Each downloader independently aborts its download after a certain time, which is exponentially distributed with mean 1/θ.

**γ** the rate at which seeds leave the system

**η** indicates the effectiveness of the file sharing

η takes values in [0, 1]

## Peer Evolution:

- In P2P file sharing, the number of peers in the system is an important factor in determining network performance.
- Therefore, it is useful to study how the number of peers evolves as a function of the request arrival rate, the peer departure rate, the uploading/downloading bandwidth of each peer, etc.

## Scalability:

- The network performance to not deteriorate, and preferably to actually improve, as the size of the network increases.
- Network performance can be measured by the average file downloading time and the size of the network can be characterized by the number of peers, the arrival rate of peers, etc.

## File Sharing Efficiency:

- Peers have different uploading/downloading bandwidths.
- A file may be broken into smaller pieces and the pieces may be distributed at random among the peers in the network.
- To efficiently download the file, it is important to design the file-sharing protocol such that each peer is matched with others who have the pieces of the file that it needs and further, to ensure that the downloading bandwidth of each peer is fully utilized.

## Incentive Mechanism:

- The objective of the incentive mechanism is to encourage users to contribute. There are three algorithms in BitTorrent which are intended to **discourage free-riding.** They encourage users to upload and contribute.
    - **Peer Selection Algorithm**
      Four principal mechanisms:
        - tit-for-tat
        - optimistic unchoking
        - upload only
        - anti-snubbing
    - **Peer Strategy**
    - **Piece Selection Strategy**
        - Strict Priority
        - Endgame Mode
        - Rarest First
        - Random First Piece
    - **Nash Equilibrium Point:**
      Each peer chooses its physical uploading bandwidth to be equal to the actual uploading bandwidth.

## Steady State Performance:

To study the system in steady state, we let :

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \frac{\mathrm{d}y(t)}{\mathrm{d}t} = 0$$

## Little's law:

**L = λW**

L – the average number of items in a queuing system

λ – the average number of items arriving at the system per unit of time

W – the average waiting time an item spends in a queuing system

## Markov chain modelling:

- A Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.
- We numerically computed the stationary distribution for this Markov chain to find mean, number of jobs, servers, and delay for t.

## Local Stability:

- We implicitly assumed that the system is stable and will reach its equilibrium. We study the stability of the fluid model around the equilibrium.

## Characterizing Variability:

- Characterization of the variance of x(t) and y(t) around their equilibrium values, using a Gaussian approximation.
- Ornstein-Uhlenbeck process
- Switched linear systems

## Effect of Different parameters on Network Performance:

We obtain the expressions for the average number of seeds, the average number of downloaders, and the average downloading time as functions of the peer arrival rate, downloader leaving rate, seed leaving rate, uploading bandwidth, etc, which explicitly give us insight on how the network performance is affected by different parameters.