Peer to Peer Distributed File Sharing System

1 Introduction

This report outlines the implementation logic of a Peer-to-Peer file-sharing system, focusing on the communication between the tracker and clients during file downloads. The system involves clients uploading and downloading files in pieces, and a tracker maintaining metadata about these files and peers. The report highlights the core functionalities of the system and how they are implemented.

2 Tracker and Client Communication

The P2P system consists of two primary components:

- **Tracker**: A centralized server that maintains metadata about peers, their IP addresses, files they have, and piece wise availability of files among the peers. This acts like a server in the whole system.
- **Client**: Peers that upload or download file pieces and interact with the tracker to get metadata of the file they want. It acts both as a server and a client. Here, peer-to-peer connection is needed.

In our system:

- **Upload File**: When a client uploads a file, it doesn't send the entire file to the tracker. Instead, it sends metadata, such as its IP, port, and the file hash. The tracker uses this metadata to allow other peers to discover this client.
- **Download File**: A client fetches metadata from the tracker about a file it wants to download and starts requesting file pieces from other peers.

3 Implementation Logic

3.1 Download Logic

The download process starts when a client requests a file from the tracker. The client receives metadata about which peers hold the pieces of the file. Based on the rarest-first algorithm, the client prioritizes downloading rarer pieces first so that they won't get missed over time.

Key components of the implementation:

Piece Selection Logic: Rarest-first algorithm was implemented by sorting file pieces based on the number of peers that hold each piece. The pieces with the fewest peers are downloaded first to ensure that rarer pieces are propagated quickly in the network.

Peer Selection: After sorting the pieces by rarity, the system selects a peer at random from the available peers for that piece. A round-robin way of selection is used for them.

3.2 Download Thread

Each file piece download happens in a separate thread. This ensures parallel downloads and improved efficiency. The download thread interacts with the selected peer and requests the specific file piece.

3.3 Reporting Back to Tracker

Once a piece is successfully downloaded, the client needs to inform the tracker. This keeps the tracker updated about which clients have which pieces, allowing it to provide accurate metadata to others who need to download the file.

4 Tracker Features

4.1 Listening for Client Connections

The tracker runs a listening thread to accept connections from clients. When a client connects, the tracker processes the client's request (upload, download, etc.) and responds accordingly.

5 Rarest-First Algorithm

The rarest-first algorithm ensures that clients prioritize downloading the least common pieces of the file first. This is crucial for maximizing the availability of all pieces in the network and preventing bottlenecks where certain file pieces are not shared enough.

5.1 Steps Involved in Rarest-First:

1. Track how many peers hold each piece of the file.
2. Sort the pieces based on availability (rarest pieces first).
3. Download the rarest piece from one of the peers holding it.

This approach increases the likelihood that all peers will eventually get the entire file because it spreads rare pieces across more peers early in the download process.

6 Conclusion

The P2P file-sharing system incorporates key components such as parallel downloads, tracker communication, and rarest-first piece selection which makes it better.

7 References

- Peer-to-peer networking with BitTorrent
- Incentives Build Robustness in BitTorrent