Understand the *Unfairness* of BitTorrent

Zengbin Zhang†, Yao Li†, Yang Chen†, Pei Cao‡, Beixing Deng†, Xing Li†
†Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
Email: {zzb02, yao-li06, chenyang04}@mails.tsinghua.edu.cn
‡Department of Computer Science, Stanford University, CA 94305
Email: cao@cs.stanford.edu

ABSTRACT

BitTorrent (BT) is the most popular P2P file-sharing application. Its tit-for-tat mechanism aims to guarantee the efficiency and fairness of sharing. However, while BT's download efficiency has been proven, we find that the current protocol suffers seriously from *unfairness*, in the sense that certain peers will always serve as *Super Peer*s. In this paper, we report on experiments conducted to pinpoint the cause of unfairness. The results indicate that the occurrence of *Super Peer* has a strong correlation with the bandwidth between the initial seed and the peer, and a weak correlation with the start time of the peer.

Keywords

bittorrent, unfairness, super peers, bandwidth, start time

1. INTRODUCTION

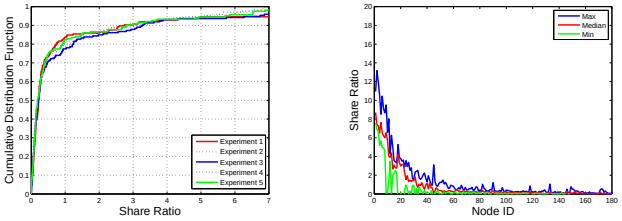
Today, BitTorrent accounts for nearly 40 percent of Internet traffic. Plenty of studies have investigated its performance. Most focus on download efficiency and service availability of BitTorrent, while little attention has been paid on fairness, i.e. whether all peers contribute equally. [1] finds through theoretical analysis that tit-for-tat is not effective in preventing unfairness. [2] proposes three modifications of the protocol to improve fairness; however, the results are mainly drawn from simulations.

With the emergence of usage-based charging by service providers, it is important to quantify and understand unfairness in BitTorrent. In this paper, we first conduct experiments on Planetlab to measure the severity of unfairness. We then attempt to find the causes of unfairness through experiments designed to measure the correlation between Super Peers and a number of features. We find that the bandwidth between the initial seed and a peer is the most significant factor in determining if the peer becomes Super Peer.

2. UNFAIRNESS IN BITTORRENT

2.1 Experimental Setup

We conduct experiments on Planetlab to evaluate the severity of unfairness. Around the globe, 180 Planetlab nodes are randomly chosen to run experiments. planetlab-3.cs.princeton.edu is the Tracker, and planetlab3.csail.mit.edu is the Initial Seed with upload rate limited to 400KB/s. There are no limits on upload rate and download rate for each peer. All peers behave selfishly, stopping immediately when downloads complete.



(a) Cumulative Distribution Function of Share Ratios (b) Share Ratio Distribution of Each Peer

Figure 1: Share Ratio

2.2 Severity of Unfairness

To measure fairness, we use a common metric called Share Ratio, defined as the ratio of bytes uploaded to bytes downloaded. We run 5 experiments, starting peers in a random order, then collecting the Share Ratio of each peer.

Fig.1(a) shows the cumulative distribution of each node's Share Ratio averaged over five experiments. In each experiment, only a few nodes contribute heavily, while others share little. About 10 peers have an average Share Ratio of around 8, which means that nearly 80 peers are served by these 10 peers. In Fig.1(b), we plot the max, median and min value of Share Ratio of each peer in 5 experiments, from which it can be seen that, even when we randomize the starting sequence in each experiment, the same small group of nodes remain the main contributor. We call these nodes *Super Peer*s.

The presence of Super Peers is surprising, especially considering the fact that the Super Peers behave selfishly too, that is, they exit as soon as their downloads complete.

3. WHAT CAUSES UNFAIRNESS

What makes a peer a Super Peer? The cause is unlikely to be platform or kernel since all Planetlab nodes are configured similarly. Hence, we look at the networking aspects of the peers.

3.1 Methodology

We use experiments to examine the relationship between Share Ratio and the following features.

Feature 1: Peer Bandwidth to Seed. We measure the network bandwidth between the Initial Seed and each peer 3

times, and compare the sequence of the average bandwidth with the sequence of the median Share Ratio.

Feature 2. Link Distance to Neighbors. The link distances (RTTs) of each peer to its neighbors are measured by Scube. The distributions are compared among the peers, especially between Super Peers and peers with low share ratio.

Feature 3. Link Bandwidth to Neighbors. The same as *Feature 2*, except that we measure the network bandwidth instead of RTT.

Feature 4: Start Time. We randomly change the start time for each peer. We perform 3 sets of experiments. In the 1st set, all peers start at the same time. In the 2nd set, peer start time is chosen randomly between 0 and half of the average download time. In the 3rd set, start time is between 0 and the average download time. We run 5 experiments in each set.

3.2 Experiment Results

Feature 1: Peer Bandwidth to Seed has the strongest relationship to Share Ratio, shown in fig.2. We use Algorithm 1 to calculate the *correlation coefficient* of the vector of Share Ratio and the vector of Bandwidth to Seed. The ratio represents how close the relationship is, with 1 meaning completely the same and 0 meaning no relationship. The result for *Feature 1* is 0.55.

Algorithm 1 Correlation Coefficient $r(\vec{x}, \vec{y})$

$n \leftarrow$ Number of Peers
$\vec{x} = [x_1, x_2, \ldots, x_n]$, $\vec{y} = [y_1, y_2, \ldots, y_n]$
$\vec{x} \leftarrow normalize(\vec{x})$, $\vec{y} \leftarrow normalize(\vec{y})$
$r(\vec{x}, \vec{y}) \leftarrow \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$
$r(\vec{x}, \vec{y}) \in [0, 1]$

We believe this is due to the fact that, when the seed in Bittorrent uses tit-for-tat to select peers to upload data, its selection criteria is strictly the download rate of the peer. As a result, peers with a high bandwidth to seed are always preferred and act as the proxy of the seed, thus contributing much more than others.

We find that *Feature 4: Start Time* also affects the formation of Super Peers. In fig.3, we calculate the average *correlation coefficient* of the start time sequence and the share ratio sequence under 3 scenarios as specified in Section 3.1. It is clear that, as the difference in start time increases, the correlation between Start Time and Share Ratio increases, though the correlation is not as significant as *Feature 1*. With an early start time, thus with little competition from others, a peer can get data easily from the seed. As a result, it tends to share much more and become a Super Peer, even though it stops when its own download completes.

We find little correlation between share ratio and *Feature 2* or *Feature 3*. Due to space limitations we do not present the results here.

4. CONCLUSION AND FUTURE WORK

It's clear that, despite tit-for-tat, some peers do much more work than others in BitTorrent, especially peers that can download from seed fast and join the torrent early. We plan to refine the BitTorrent protocol to reduce the unfairness.

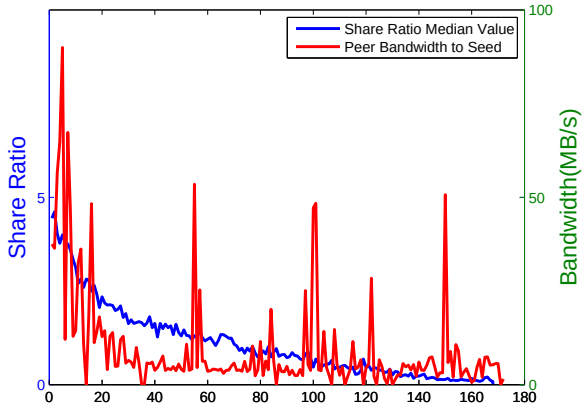


Figure 2: Feature 1. Peer Bandwidth to Seed

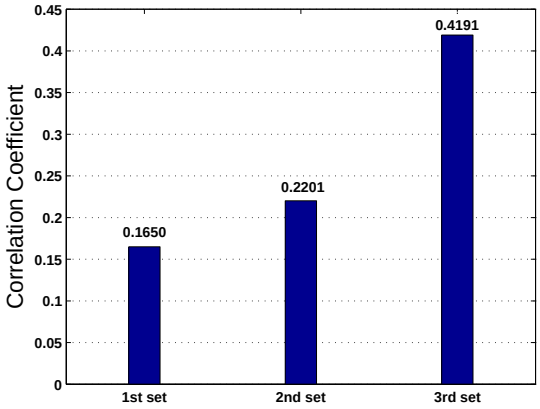


Figure 3: Feature 4. Start Time

5. ACKNOWLEDGMENT

This work is supported by the Project Research on Method and Security of Network Coordinate Based Superspace-Subspace Embedding supported by NSFC (No.60850003), the National Basic Research Program of China (No.2007CB310806) and the National High Technology Development Program of China (No.2007AA010306).

6. REFERENCES

[1] A. R. Bharambe, C. Herley, and V. Padmanabhan. Analyzing and improving a bittorrent networkś performance mechanisms. In *INFOCOM*, 2006.

[2] R. Thommes and M. Coates. Bittorrent fairness: Analysis and improvements. In *WITSP*, 2005.