Influences on Cooperation in BitTorrent Communities

Nazareno Andrade
Universidade Federal de Campina Grande
nazareno@dsc.ufcg.edu.br

Miranda Mowbray
HP Laboratories Bristol
miranda_mowbray@hp.com

Aliandro Lima
Universidade Federal de Campina Grande
aliandro@dsc.ufcg.edu.br

Gustavo Wagner
Universidade Federal de Campina Grande
gustavo@dsc.ufcg.edu.br

Matei Ripeanu
University of Chicago
matei@cs.uchicago.edu

ABSTRACT

We collect BitTorrent usage data across multiple file-sharing communities and analyze the factors that affect users' cooperative behavior. We find evidence that the design of the BitTorrent protocol results in increased cooperative behavior over other P2P protocols used to share similar content (e.g. Gnutella). We also investigate two additional community-specific mechanisms that foster even more cooperation.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: retrieval models; C.4 [**Performance of Systems**]: performance attributes; H.1.2 [**User/Machine Systems**]: human factors

General Terms

Measurement

Keywords

BitTorrent, P2P, cooperation

1. INTRODUCTION

All collaborative computing systems, including P2P file-sharing systems, potentially face the problem of freeriding: that is, users or peers that consume resources of the system without contributing anything in return. Adar et al. [1] found in 2000 that 66% of Gnutella users were freeriders: users that did not make any files available for download to other users. In 2005, Hughes et al. [9] classified 85% of Gnutella users as freeriders. These and other studies [15] demonstrate a lack of cooperation in traditional P2P file-sharing systems.

BitTorrent [3] is a peer-to-peer (P2P) file-distribution tool that employs a tit-for-tat incentive mechanism to reduce freeriding and increase user cooperation. It has proved extremely popular: CacheLogic estimates that BitTorrent generated about 30% of all US Internet traffic in June 2004 [5].

BitTorrent works as follows: the original file distributor publishes details of the file on a web server, and creates a *tracker* that allows peers interested in the file to find each other. To download the file, peers access the tracker and join the torrent (in BitTorrent lingo, a *torrent* is a group of peers connected to the same tracker). The file is divided into chunks. As a peer downloads chunks of the file, it also uploads, to other peers in the torrent, chunks that it has previously downloaded. Thus, the burden of bandwidth consumption is moved from the original content distributor to all peers in the torrent.

A distinguishing feature of BitTorrent is its incentive for cooperation. Peers reciprocate: a peer is most likely to upload (i.e. serve content) to those peers that have recently uploaded to it [6]. This gives an incentive for peers to cooperate by uploading to other peers while they download.

In this paper we use logs collected from five different file-sharing communities to analyze (i) the cooperative behavior induced by the BitTorrent protocol, and (ii) the impact of content type and community-wide policies on the level of cooperation in these communities. We find that, although there are circumstances in which the BitTorrent incentive mechanism does not work, the freeriding levels in these communities are low. However, in some file-sharing communities, the cooperation levels are not as high as are desired by community administrators. In these cases, additional community-specific policies can successfully boost cooperation.

The rest of this paper is structured as follows. Section 2 presents related work and the metrics used to estimate cooperation levels. BitTorrent community structure and the mechanisms that leverage it are presented in Section 3. Section 4 outlines our methodology and Section 5 presents our results. We summarize in Section 6.

2. RELATED WORK AND METRICS

A number of previous studies have analyzed existing BitTorrent deployments or used analytical modeling [14] and simulations [2] to characterize the properties of the BitTorrent protocol and to improve its performance.

Izal et al. [10] use tracker logs and an instrumented client to analyze the behavior of 180,000 peers in a torrent distributing Linux RH9 over five months. The download and upload rates of their instrumented peer are positively correlated, illustrating the effectiveness of the incentive mechanism.

Pouwelse et al. [13] monitor a very popular BitTorrent site and report on its observed characteristics: data availability and integrity, flashcrowd handling, and download performance.

Qiu and Srikant [14] use a simple fluid model to study steady state performance for a BitTorrent-like system. They prove that, under certain conditions, a Nash equilibrium exists, and derive analytical upper bounds for freeriding.

However, to the best of our knowledge, no previous study provides empirical evidence on the effectiveness of the BitTorrent incentive mechanism in promoting cooperative user behavior in real deployments, or compares several BitTorrent communities using the same cooperation metrics. To shed light on this question, we define three metrics for cooperation: *freeriding*, *seeding* and *sharing ratios*, and evaluate them in several BitTorrent communities.

2.1 Freeriding

A *freerider* downloads but does not upload any data. This may happen when the user specially configures or modifies his client software. There may be multiple motivations to attempt to freeride: a user might try to avoid the bandwidth cost of uploading [7], or, in the case of copyright-infringing content, the user may see uploading as more immoral or more risky. Additionally, firewalls impact peers' ability to contribute. If two peers are behind firewalls, they cannot exchange chunks of the file.

The *freeriding ratio* for a P2P file-sharing community is the percentage of peers that are freeriders. The low freeriding ratios we observe in some existing BitTorrent communities indicates that BitTorrent is successful in boosting cooperation.

2.2 Seeding

A *seeder* is a peer that has finished downloading but is still connected to the torrent, and might upload to other peers. A torrent benefits from seeders, as they generally increase content availability and download rates. BitTorrent sites and client software exhort users to allow their peers to seed; however, no incentive for seeding is present in the protocol.

We investigate typical seeding behavior over time. We also look differences in *seeding ratio*, i.e., the ratio of seeders to participating peers in a torrent, between different BitTorrent communities, and suggest reasons for the significant differences that we find.

2.3 Sharing ratio

The *sharing ratio* of a peer is the total amount of data the peer has uploaded, divided by the total amount downloaded. By analogy, we define the sharing ratio of an entire torrent at a specific moment to be the total amount uploaded divided the total amount downloaded by the peers active in the torrent at that moment.

The sharing ratio is related to the two metrics above: a freerider is a peer whose sharing ratio is equal to zero, and a seeder uploading content will thereby increase its sharing ratio.

Most popular BitTorrent client implementations allow users to set the maximum upload rate. None of the ten most popular clients we investigated allows setting the maximum upload rate to zero (without modifying the source code) but they do allow setting a low rate. As a result, a user who wishes to freeride may fail to do so, but succeed to upload only a small amount of data relative to the amount downloaded. In our data analysis we therefore investigate the frequency of peers with a low sharing ratio.

3. BITTORRENT COMMUNITIES

BitTorrent files are usually published through websites that consist of listings of torrents with HTML links to the trackers used for joining them [12]. An important distinction between BitTorrent and other P2P systems is the presence of these centralized components used for discovery and accounting.

This centralized architecture has several possible drawbacks: a single point of failure (illustrated for example by the failures in *suprnova.org* documented by Pouwelse et al. [13]), potential bottlenecks, and higher operational costs to maintain the centralized components.

However, the architecture allows implementation, at the site level, of new functionalities to boost user cooperation. We discuss two examples in the rest of this section.

3.1 Sharing-ratio enforcement

We shall see in Section 5 that the BitTorrent protocol does appear to successfully encourage a relatively high level of cooperation between peers. However, some BitTorrent site operators feel that additional mechanisms to enforce even more cooperation are beneficial. Thus, several BitTorrent communities, including *easytree.org*, *empornium.us*, and *pwtorrents.net*, periodically enforce that peers are above a minimum sharing ratio: these sites keep a long-term history of user cooperation, and prevent peers below a certain sharing-ratio threshold from gaining access to new content. (In some cases this decision mechanism also takes peer "age" into account.)

This mechanism is generally implemented by requiring users first to register and login to the central site, and linking their identity with their download activity. Thus, user anonymity is sacrificed.

Since a peer can efficiently increase its sharing ratio by uploading content as a seeder, the mechanism provides an indirect incentive for seeding as well as a direct incentive not to freeride.

We study how sharing-ratio enforcement affects user behavior at *easytree.org*. We are motivated by the report by the site administrators that the level of cooperation increased when sharing ratio enforcement was introduced [11].

The Maze P2P file-sharing mechanism employs a centralized mechanism similar to the sharing-ratio enforcement: points are awarded for file uploads. Yang et al. [18] show that Maze users circumvent the incentive mechanism by leaving and re-entering as new users. The administrators of *easytree.org* report that they have successfully limited this problem by imposing restrictions on the creation of new accounts, so that a user who tries to reenter the system with a new identity may have to wait for a long time.

3.2 Broadcatching

Several BitTorrent websites use RSS feeds [4] to advertise newly published files. Broadcatching is the use of BitTorrent clients to automatically download files advertised through RSS feeds. For example, a user may subscribe to an RSS feed of a site that publishes past episodes of TV series, and state interest in any new episode from a particular series. Whenever the RSS feed announces matching content, the client will download it automatically. The centralized website makes data integrity checking and standard formats for file naming easier to enforce, so that users can trust that the files advertised will be the ones delivered.

In the time between a client finishing a download and the user checking to see whether new files have arrived, the client remains connected as a seeder. So broadcatching may result in more cooperation, because users may keep their peers running for longer as a side-effect of broadcatching.

4. METHODOLOGY

We collected data from five BitTorrent communities: *bt.etree.org*, *piratebay.org*, *torrentportal.com*, *easytree.org*, and *btefnet.net*. From now on we refer to these as *etree*, *piratebay*, *torrentportal*, *easytree* and *btefnet*, respectively. Table 1 summarizes the main characteristics of these communities.

	Content	# torrents	# peers
etree	Music	567	4,492
easytree	Music	2,586	25,687
piratebay	Films, music etc	13,054	320,900
torrent-portal	Films, music etc	10,115	357,428
btefnet	TV episodes	476	78,897

Table 1: Characteristics of the communities

We collect data by crawling the public torrent report pages for the site. Each crawling provides a snapshot of the community at a given moment. Since we obtain data from a large number of torrents of different ages, we expect to have a representative sample from the different stages of a torrent's life.

We collect data on all the torrents that are active at the time of sampling, have at least three peers with upload or download greater than zero, and have at least one seeder, so that there is at least some sharing going on.

For each torrent at each site we collect its corresponding file size, age in days, number of participating peers, and seeding ratio. In addition, for *etree* and *easytree* we are able to collect data about the state of each participating peer: the amount downloaded and uploaded, whether the peer is a seeder, and whether it is *connectable*, that is, whether the tracker is able to open a connection to it. The *etree* data is public, while the *easytree* data was provided by the system administrators, and is analyzed here for the first time.

Some of the *etree* torrents have zero download, resulting in an infinite sharing ratio. Therefore instead of investigating the factors influencing the sharing ratio directly, we investigate those influencing torrent rank, where torrents are ranked according to their sharing ratio (torrents with high ratios having high rank).

Where we say that there is a positive (or negative) correlation between two variables, we mean that the Pearson correlation coefficient is positive (or negative) and is significant at the 0.01 significance level using a one-tailed t-test [8].

5. RESULTS

In this section we look at relative download times for freeriders and non freeriders, measure the amount of freeriding and low-sharing in *etree* and *easytree*, and investigate the factors affecting the seeding ratio in the five sites.

5.1 Relative download times

In torrents with a relatively low number of seeders, BitTorrent is successful in penalizing freeriding, in effect by increasing the download times of peers that freeride. However, in torrents where seeders are plentiful, i.e., torrents with high seeding ratios, freeriders may download faster than collaborating peers (Figure 1). This might be a direct manifestation of the cost of cooperation [7]: TCP acknowledgment packets compete with (and slow down) incoming data streams. Figure 1 plots the ratio between download time experienced by a freerider and that experienced by a collaborating peer varies with the torrent seeding ratio. The experiment was conducted in 21 torrents in *etree*, *btefnet*, *piratebay* and *suprnova.org*, in October 2004 and March 2005, using two "snark" clients [16], one of which was modified to make it freeride. Data on the seeding ratios of the torrents was provided by the torrent trackers.

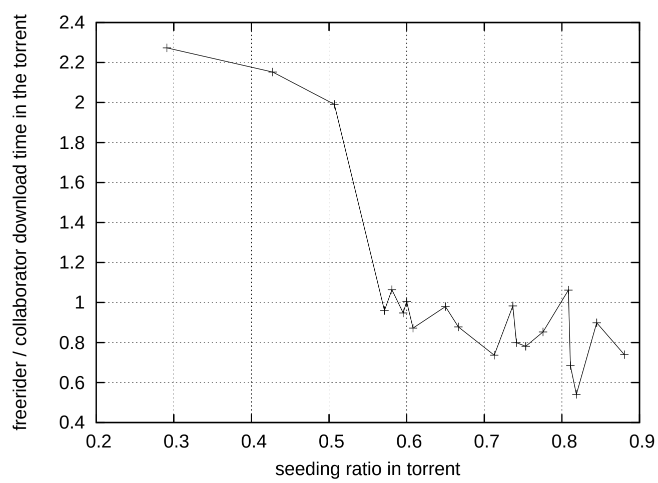

Figure 1: Ratio between download time experienced by a freerider and that experienced by a collaborating peer as the seeding ratio varies.

5.2 Freeriding and sharing ratio

Fewer than 6% of the live peers in our sample from *etree* have not uploaded anything. Of these, only 19% were connectable, whereas nearly half of all live peers were connectable. The median age of the peers that have not uploaded anything was 3.3 hours, whereas the median age of all live peers was 10.6 hours. We sampled *etree* peers again

3 days later, 6 days later, and several weeks later, and did not find any dramatic changes in these values.

At *easytree*, only 5% of the live peers uploaded nothing. Of these, 41% were connectable, whereas 60% of all live peers were connectable. There were fewer low-sharing peers, 24%, than at *etree* (28%). The age of *easytree* peers is only recorded to the day; however, 72% of peers in this group had an age of one day or less, compared to 65% of all live peers. (The threshold used to define low sharing ratio is 0.25, a value decided after extended discussion and a poll on *easytree*'s mailing list [11]).

We believe that these slightly higher levels of cooperation at *easytree* are explained by the use of sharing-ratio enforcement against low-sharing peers. As we will show later, *easytree* also has significantly higher levels of seeding.

At both sites, the group of peers that have not uploaded anything include disproportionately many peers of low age. If a peer has not downloaded any chunks needed by other peers in the torrent, then it will be unable to upload data to them. Thus peers that have not uploaded data are not necessarily trying to freeride: they may be willing to upload but unable to at present. Indeed, some peers that had uploaded nothing at the time of our first *etree* sample, uploaded data later. Also, some peers may be low-sharing as a result of asymmetric bandwidth links.

Freeriders and other low-sharing peers are much rarer than freeriders are in other P2P communities (for example Gnutella [1, 9, 15]). It appears therefore that the design of the BitTorrent protocol does result in increased cooperative behavior.

5.3 Peers' seeding behavior

In this section we investigate peers' seeding behavior over time as reflected in *etree* data. In the next section we focus on the relationship between seeding rates and various torrent characteristics across different sites.

To investigate typical seeding behavior, we have collected hourly snapshots of seeders at *etree* for a 10-day period in March 2005. A peer that appears in a snapshot as a seeder is considered to continue seeding until the first snapshot where it is no longer reported as participating in the torrent. The results are shown in Figure 2. There is a diurnal rhythm in the number of seeders, which lies between 2500 and 3700. A majority of seeders do not remain seeding for more than a day, so the collection of peers that are seeding changes over time although the total number of seeders remains bounded.

5.4 Correlations of torrent characteristics

We now look at torrents, instead of peers.

For *easytree* and *etree*, the two sites for which we have data on sharing ratios, we rank torrents according to their sharing ratios. We find that torrent rank is:

- *positively correlated* with torrent seeding ratio. This is not surprising since seeders upload without downloading.

- *positively correlated* with the number of peers in the torrent. This correlation is not explained by increased seeding ratio for large torrents, since there is not a positive correlation between the seeding ratio and number of peers at any of the five sites. A possible contributory factor is that a firewalled peer can only cooperate with peers in the same torrent that are not behind firewalls themselves. The larger the number of peers in a torrent, the more likely it is that such a peer exists.

- *not correlated* with the age of the torrent or the size of the file shared.

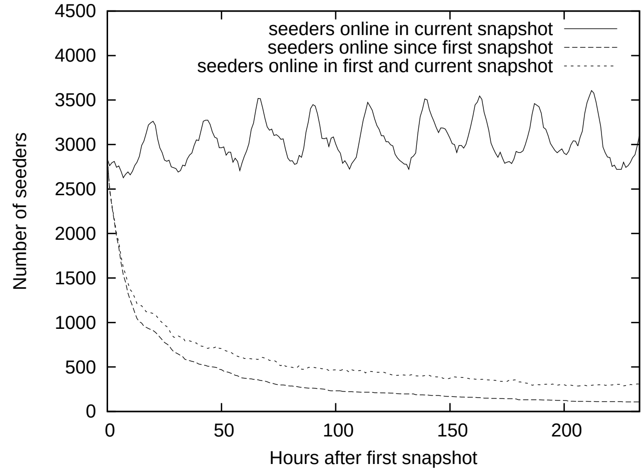

Figure 2: Seeding over time in etree.

At all five sites, torrent seeding ratio is negatively correlated with the size of the file (i.e., increased seeding ratio for torrents sharing small files). A possible explanation is that the amount of time peers remain connected to the torrent after they finish downloading is independent of their download time. Since small files download faster, the time that a peer spends as a seeder is relatively longer for torrents sharing small files than for torrents sharing large files.

5.5 Seeding: cross-site comparisons

In addition to comparing seeding ratios across torrents of a single site, we can compare seeding ratios across different sites, to indirectly estimate the relative levels of cooperation in different BitTorrent communities. Figure 3 shows a cumulative graph of seeding for each of the five sites.

As expected, among the sites that do not use additional mechanisms to increase cooperation (*etree*, *piratebay* and *torrentportal*), the site that enables distribution of only legal content (i.e., *etree*) has the largest seeding rates.

Btefnet torrents also have significantly more seeding than those in *piratebay* and *torrentportal*. The use of broadcatching is a possible explanation.

Easytree torrents have significantly more seeding than those at any other sites. We attribute this to the sharing-ratio enforcement.

These differences in seeding between sites are all significant at the 0.01 significance level. We have verified using partial regressions that they are not explained by differences in torrent age or file size.

The social characteristics of different communities strongly influence sharing behavior. The moderator of *easytree* (a site that enables distribution of bootleg recordings [17] and has systems in place to prevent the distribution of commercially-released music) attributes the relatively high amount of cooperation to the sharing culture among offline bootleg traders. When there was a large influx of new users who did

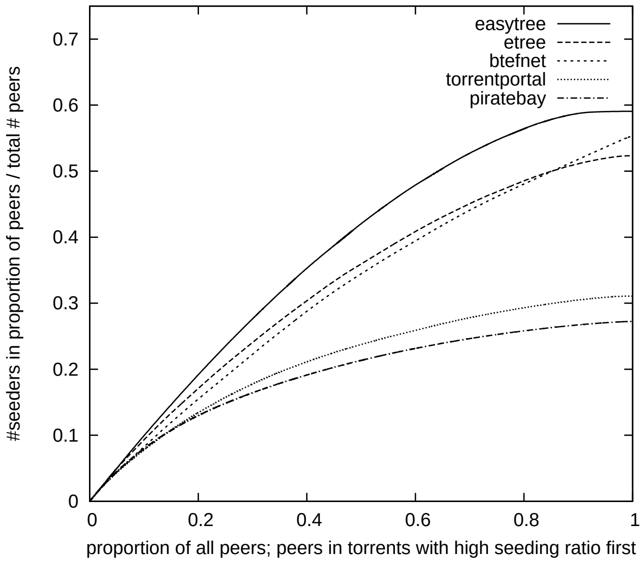

Figure 3: Cumulative seeding ratio for all sites.

not have this background, the sharing ratio went down, and site administrators successfully introduced sharing-ratio enforcement to reverse its decline [11].

6. SUMMARY

Data collected across a set of BitTorrent communities supports the claim that the incentives for cooperation employed by the BitTorrent protocol do succeed in discouraging freeriding. These incentives however do not lead to the cooperation levels desired by some file-sharing communities. In these cases additional community-specific mechanisms to boost the levels of cooperation are successfully employed.

Our comparison across sites shows significant variations in seeding levels, and suggests that social as well as economic characteristics play a role in determining the amount of cooperation.

At all the sites we measure, we find that the seeding ratio is higher in torrents with small file sizes and, at all but one of the sites, is higher in younger torrents. At the two sites where we are able to rank torrents according to their sharing ratios, torrent rank is positively correlated with seeding ratio and with the number of peers in the torrent.

We also find that in torrents with a large number of seeders, the BitTorrent tit-for-tat mechanism may not succeed in producing a disincentive for freeriding: in such torrents, freeriders may actually experience faster download times than cooperating peers.

Acknowledgments

The authors thank the *easytree.org* administrators Jupiter and Erwe for their helpful discussions and data. This work was partially developed in collaboration with HP Brazil R&D. Matei Ripeanu was supported in part by a grant from IBM Research.

7. REFERENCES

[1] E. Adar and B. Huberman. Free Riding on Gnutella. *First Monday*, 5(10), October 2000.

[2] A. Bharambe, V. Padmanabhan, and C. Herley. Understanding and Deconstructing BitTorrent Performance. In *ACM SIGMETRICS*, June 2005. http://www.research.microsoft.com/˜padmanab/papers/msr-tr-2005-03.ps.

[3] BitTorrent, Inc. BitTorrent web site, 2001–2005. http://www.bittorrent.com.

[4] J. Bond. RSS FAQ. http://www.voidstar.com/module.php?id=129&mod=book&op=feed.

[5] CacheLogic. The true picture of peer-to-peer filesharing. http://www.cachelogic.com/research, July 2004.

[6] B. Cohen. Incentives Build Robustness in BitTorrent. In *Proc. Workshop on Economics of Peer-to-Peer Systems*, Berkeley CA, USA, June 2003.

[7] M. Feldman, K. Lai, J. Chuang, and I. Stoica. Quantifying disincentives in peer-to-peer networks. In *Workshop on Economics of Peer-to-Peer Systems*, June 2003.

[8] P. R. Hinton. *Statistics Explained*. Routledge, 2004.

[9] D. Hughes, G. Coulson, and J. Walkerdine. Freeriding on Gnutella Revisited: the Bell Tolls? *IEEE Distributed Systems Online*, 6(6), 2005. http://dsonline.computer.org/portal/pages/dsonline/0506/o6001.html.

[10] M. Izal, G. Urvoy-Keller, E. W. Biersack, P. Felber, A. A. Hamra, and L. Garcés-Erice. Dissecting BitTorrent: Five Months in a Torrent's Lifetime. In *Proc. Passive and Active Measurements (PAM2004)*, Antibes Juan-les-Pins, France, April 2004.

[11] Jupiter (Easytree Chief Moderator). Personal communication, March 2005.

[12] Monkey Methods research group. Monkeyin' around: Is BitTorrent dead? http://monkeymethods.org/pubs/is-bittorrent-dead-centralization-analysis.htm, Jan. 2005.

[13] J. Pouwelse, P. Garbacki, D. Epema, and H. Sips. The Bittorrent P2P File-sharing System: Measurements and Analysis. In *4th International Workshop on Peer-to-Peer Systems (IPTPS'05)*, Feb. 2005. http://www.st.ewi.tudelft.nl/˜pouwelse/Bittorrent_Measurements_6pages.pdf.

[14] D. Qiu and R. Srikant. Modeling and performance analysis of BitTorrent-like peer-to-peer networks. *ACM SIGCOMM Computer Communications Review*, 34(4):367–367, Oct. 2004. http://www.ifp.uiuc.edu/˜dqiu/paper/ bittorrent.pdf.

[15] S. Saroiu, P. K. Gummadi, and S. D. Gribble. A Measurement Study of Peer-to-Peer File Sharing Systems. In *Proc. Multimedia Computing and Networking 2002 (MMCN '02)*, San Jose CA, USA, January 2002. http://www.cs.washington.edu/homes/gribble/papers/mmcn.pdf.

[16] M. J. Wielaard. The Hunting of the Snark Project - BitTorrent Application Suite. http://www.klomp.org/snark/, 2003.

[17] Wikipedia. Bootleg recording. http://en.wikipedia.org/wiki/Bootleg_recording, 2005.

[18] M. Yang, Z. Zhang, X. Li, and Y. Dai. An empirical study of free-riding behavior in the Maze P2P file-sharing system. In *IPTPS '05*, Feb. 2005.