Public and private BitTorrent communities: A measurement study

M. Meulpolder, L. D'Acunto, M. Capotă, M. Wojciechowski, J.A. Pouwelse, D.H.J. Epema, H.J. Sips

Parallel and Distributed Systems Group
Department of Computer Science, Delft University of Technology, the Netherlands
m.meulpolder@tudelft.nl

***Abstract*—BitTorrent communities, both public and private, are immensely popular in the Internet, with tens of millions of users simultaneously active at any given moment. Public and private BitTorrent communities are managed in different ways – for instance, some private communities enforce sharing ratios, have strict rules for content management, have a certain level of community oversight, and maintain a strong sense of exclusiveness. In this paper, we present the results of extensive measurements of more than half a million peers in five communities, ranging from highly popular and well-known public communities to elite private communities that can only be joined by invitation. We observe that the performance experienced by downloaders in the private communities is by far superior to the performance in the public communities, and we observe significant differences in connectability, seeder/leecher ratio, and seeding duration. Based on our results, we conjecture that when effective ratio enforcement mechanisms are in place, BitTorrent's tit-for-tat mechanism is hardly influential anymore. Our multi-community, multi-swarm measurements are significantly broader and more extensive than any earlier measurement study on BitTorrent.**

I. Introduction

BitTorrent has become immensely popular over the last six years. Whereas initially only public trackers and their communities existed, the landscape of BitTorrent has become much more diverse, and nowadays includes large numbers of private communities with varying notions of membership and management mechanisms such as sharing-ratio enforcement and injection restrictions. In this paper, we present the results of extensive measurements of five important communities that are representative for the wide range of communities that exist in the Internet today. Our aim is to offer more detailed insight into the properties of content-sharing communities than available hitherto, especially with regard to the differences between public and private communities.

Currently, tens of thousands of BitTorrent-related websites offer services such as content search, forums, moderation, and account management, and are surrounded by communities of millions of users. In The Pirate Bay alone, around 21 *million* users are active at any given moment in time. Apart from the well-known public communities, there are increasing numbers of *private* communities, some of them serving a highly elite set of heavy users – sometimes even equipped with *seed boxes* that are dedicated to serving content 24 hours a day. At the extreme end of the spectrum a direct, personal invitation is the only way to gain access to the content; such invitations are hard to get, and even the slightest abuse leads to unconditional banishment.

This work was partially supported by the European Community's Seventh Framework Programme through the P2P-Next project (grant no. 216217) and the QLectives project (grant no. 231200).

The differences between public and private communities have hardly been quantified or qualified by the P2P research community. Real life measurement studies to date have been quite limited in scope; they often analyze just a few torrents [7], [13], a single community [5], or limited snapshots of multiple general communities [1]. The most extensive community experiments that have been presented so far [2] focus on torrent popularity and resource allocation in a three communities. Our measurements significantly extend the scope of previous work, and give clear new insights into the nature of the enormous traffic among BitTorrent users in the Internet.

In this paper, we provide the following contributions: (i) We present a selection of the results of our extensive measurements of 508,269 peers in 444 swarms of five BitTorrent communities (see Table I), ranging from public to highly elite. We observe download performance, connectability, seeder/leecher ratios, seeding duration, and statistics regarding the resource supply; (ii) We compare our results with our earlier measurements in 2003–2004 [10] and conclude that the download performance and seeding duration have increased significantly over the past 5 years; (iii) We conjecture that ratio enforcement mechanisms are the primary cause of the high numbers of seeders in the private communities, and in the end render BitTorrent's tit-for-tat reciprocity mechanism virtually irrelevant in such communities.

II. BitTorrent Communities

BitTorrent relies on central servers that run *trackers*, simple processes that keep track of the users that are downloading or seeding content and mainly serve to provide peers with addresses of other peers interested in the same content. By now, the Internet contains thousands of BitTorrent trackers – some of them very popular with extensive communities and websites around them. In addition to a content database, many such websites provide community-based functionality such as content rating, comments, and forums. Another very popular service that some websites frequently provide is a *web feed* (usually RSS), which allows users to easily and quickly discover newly published content. While many trackers can be used without any credentials, a number of trackers employ voluntary or obligatory *user accounts*. User accounts can be used by administrators to restrict the injection of content, to keep track of users' up- and download behavior, and to ban users that violate the tracker's policy. Some trackers maintain exclusive communities in which an account is obligatory and can be obtained only by *invitation*. Tracker policies in such communities are often very strict, and the violation of the rules leads to the cancellation of the account.

For our measurements we have made a selection of five communities, both public and private. A summary is given in Table I. Together these communities consist of millions of active users and millions of torrents. An overview of their primary management policies is as follows: (1) The Pirate Bay: this community is by far the most well-known, largest public community online. It has no download restrictions; (2) EZTV: this community is well-known and completely public, and has no download restrictions as well; (3) TV-Torrents: this community is well-known, but has closed membership which can only be obtained by invitation. The tracker uses a credit system where downloading costs 1 credit per byte, uploading to regular swarms yields 1 credit per byte, and uploading to underseeded swarms yields 1.5 credits per byte. A member with a zero balance is not allowed to download, and hence has to upload to earn credits first; (4) TorrentLeech: this community is reasonably well-known, and has closed membership which can only be obtained by invitation. Members have to seed each downloaded file up to a ratio of 0.4 (i.e., 4 bytes uploaded for every 10 bytes downloaded) or have to seed the completed download for at least 24 hours. In addition, a minimum overall ratio of 0.4 is required. Members who do not seed enough are warned, and have 5 days to regain their ratio to prevent losing their account; (5) PolishTracker: this community tries to keep its exposure to a minimum in order to maintain an 'elite' atmosphere; membership can only be obtained by invitation. The tracker has a very strong ratio-enforcement policy, where every downloaded file has to be seeded until the ratio is 1.0 or for at least 48 hours. In addition, an overall sharing ratio of 0.55 has to be maintained. Members that do not seed enough will first be warned and then lose their account.

Community (profile)	Sharing policy	m: # members u: avg # users t: # torrents (where known)
The Pirate Bay (public)	unlimited downloading	m: 4,000,000 u: 21,000,000 t: 2,200,000
EZTV (public)	unlimited downloading	u: $>$ 2,000,000 t: 5,490
TVTorrents (private)	1 credit / byte down, 1 or 1.5 credit / byte up, balance ≥ 0	t: 13,000
TorrentLeech (private)	seed each file for 24 hrs or until ratio $\geq$0.4, overall ratio $\geq$ 0.4	m: 178,000 t: 24,000
PolishTracker (private)	seed each file for 48 hrs or until ratio $\geq$ 1.0, overall ratio $\geq$ 0.55	m: 20,000 t: 5,750

TABLE I
PROPERTIES OF THE TRACKERS WE SELECTED FOR OUR EXPERIMENTS (SEPTEMBER 2009).

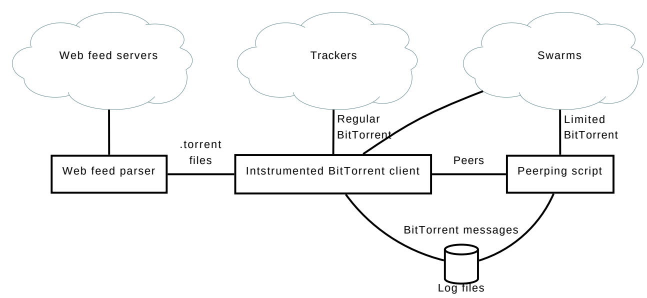

Fig. 1. The infrastructure of our large scale measurements of BitTorrent communities.

III. EXPERIMENTAL SETUP

The software infrastructure that we have employed for our measurements is depicted in Fig. 1. It consists of three main components: (1) A *web feed parser* which downloads newly published .torrent metadata files based on subscriptions to the communities' web feeds. Hence, as soon as a new piece of content is released, our software detects its presence; (2) An *instrumented BitTorrent client* which logs all the tracker communication and all the state-messages received from the peers it has connections with. We inserted an instrumented client into swarms that we discovered via the web feed parser; (3) A *peerping script* which repeatedly contacts every peer discovered by our client. The script connects to each peer every 20 seconds using BitTorrent's *Peer Wire Protocol*, performs the initial handshake, waits until it receives the *bitfield message* of the peer, and closes the connection. The received message contains information about the part of the download that the peer has completed, and enables us to induce information about download performance, seeders lifetime, seeder/leecher ratio, and connectability. Using our measurement infrastructure, we performed measurements of 508,269 peers in 444 swarms from September to December 2009 and collected over 20 million bitfields in total.

IV. MEASUREMENT RESULTS

Ratio enforcement mechanisms coupled with unique member accounts are employed by private communities to increase the number of seeders and the seeding capacity, eventually aimed at reaching a higher download performance. We therefore measured the characteristics that relate to this aim, namely: download performance, connectability, seeder/leecher ratio, seeding duration, and the fraction of data supplied by seeders.

community	measured peers	measured swarms	download speed (kbps)			avg % unconn	avg s/l ratio	seeding duration (hrs)		
			mean	median	top 10%			mean	median	top 10%
The Pirate Bay	20127	41	1037	333	>2134	47.0	2.6	11.7	1.8	>31.4
EZTV	394532	92	928	294	>1575	48.3	6.6	18.1	4.7	>52.0
TVTorrents	60900	114	3590	1362	>7692	32.5	104.5	44.1	17.9	>130.7
TorrentLeech	20874	98	4937	1030	>7166	33.9	25.4	50.4	16.8	>153.9
PolishTracker	11836	99	8625	1331	>14128	20.6	63.8	58.0	20.2	>156.0
All	*508269*	*444*	*1424*	*361*	*>2464*	*39.3*	*48.9*	*23.1*	*5.2*	*>70.4*

TABLE II
STATISTICS OF OUR RESULTS PER COMMUNITY.

A. Download performance

We measured for each community the average download speed of each discovered peer, based on the first and last bitfield messages received from it. Fig. 2 shows the CDF of the average speed per community, while Table II shows the mean, median, and maximum of the observed values. Note the log-scale on the horizontal axis of the CDF. The median download speed in the private ones is 3–5 times higher than in the public ones. The difference in mean download speed is far more extreme, suggesting that a minority of peers in the private communities has an extremely high performance. The CDF shows that at least 7% of the peers in the private communities had average speeds of 10 Mbps or higher, whereas in the public communities virtually no peers reached this average speed. Moreover, the private community with the strictest ratio enforcement (PolishTracker) shows the highest speeds.

In our earlier measurements [10], we observed that only 10% of the peers had a download speed above 520 kbps, with an average of 240 kbps. In 2005, Guo *et al.* [5] measured an average download speed of 160 kbps in a 48-day trace of 1,500 torrents. In 2006, Iosup *et al.* [6] measured a considerably higher average download speed of around 500 kbps for the top 2,000 torrents of The Pirate Bay. In our current measurements, the average download speed was around 1 Mbps in the public communities, and 3.6–8.6 Mbps in the private communities. As Table II shows, 10% of the peers in EZTV had a speed of more than 1.5 Mbps while 10% of the peers in TorrentLeech had a speed of more than 7 Mbps. Furthermore, 36–40% of the peers in the public communities and as much as 64–72% of the peers in the private communities had a download speed above 520 kbps. The average speeds in the currently measured public communities are 4 times higher than those measured in 2003–2004, while for all the peers we measured this is almost 6 times. The average speed in PolishTracker is even 36 times higher than that in any community we measured in 2003–2004. We can safely conclude that the performance has seen a significant increase over the last 5 years.

B. Connectability

Mol *et al.* [9] show that under a given fraction of unconnectable peers (e.g., peers behind a NAT or firewall), there is an upper bound to the sharing ratio these peers can sustain as a group. We would therefore expect lower

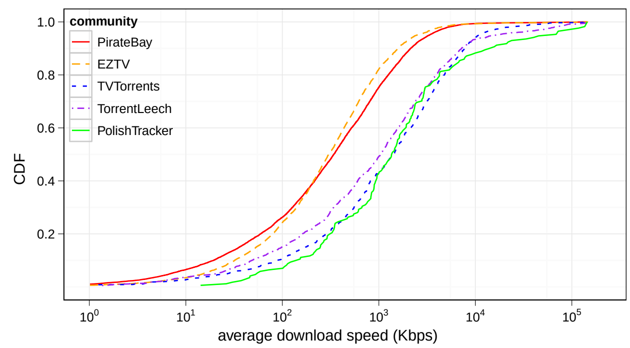

Fig. 2. The CDF of the download speed per community.

fractions of unconnectable peers in private communities. In order to investigate this effect, we measured the fraction of unconnectable peers in each of the five communities.

Fig. 3 depicts the average fraction of unconnectable peers in each community over time, where time is taken relative to the birth of the respective swarm. Table II displays the overall average fraction[1] per community. Clearly, the gap between the public communities and the private communities is considerable. At the extremes, public community EZTV has 47% unconnectable peers on average while elite community PolishTracker has only 20% unconnectable peers on average.

In our earlier measurements of 2003–2004 [10], we observed around 40% unconnectable peers, which is in line with the current overall average of 39.3%. However, the overall fractions of unconnectable peers in our measurements are lower than those reported in other work. Measurements of [9] show 66% unconnectable peers for the public community Pirate Bay, and 45% for the private community TVTorrents. Xie at al. [15] report 70% unconnectable peers for the public CoolStreaming system.

C. Seeder/leecher ratio

The seeder/leecher ratio indicates the number of seeders per leecher, and therefore gives an idea of supply vs. demand

[1] We first computed the average fraction of unconnectable peers *per swarm* over its lifetime, and then computed the overall average of the per-swarm averages.

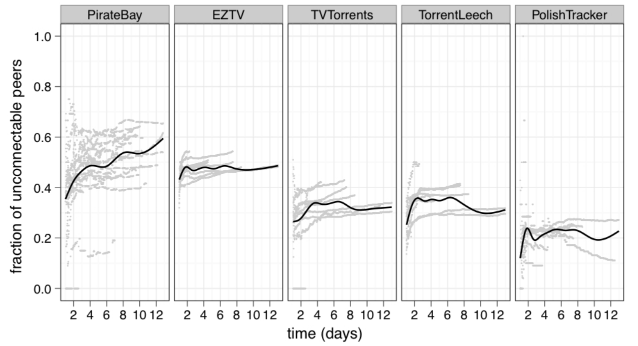

Fig. 3. The average fraction of unconnectable peers per community over time (since swarm birth). Each grey dot represents an observation of a swarm at a point in time, while the lines are a locally weighted scatterplot smoothing.

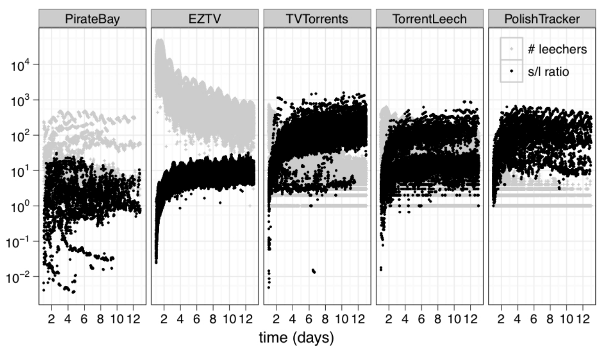

Fig. 4. The seeder/leecher ratio and number of leechers per community over time (since swarm birth). Each dot represents an observation of a swarm at a point in time.

in swarms. Fig. 4 shows observations of the seeder/leecher ratio in swarms in each of the communities. Table II gives the overall average ratio[2] per community. In TVTorrents, there are on average more than 100 seeders per leecher, with a peak observation of 1589. PolishTracker has 64 seeders per leecher, with a peak observation of 667. In such 'overseeded' swarms, it is likely that piece requests from an arriving leecher are almost immediately granted by seeders, and that the leecher can therefore saturate its download capacity quickly. Moreover, during our measurements none of the swarms in PolishTracker was *ever* observed to drop below a ratio of 1.

The public communities, however, have considerably lower seeder/leecher ratios. On average, the public communities had only 2–7 seeders per leecher. Even at peak observations, The Pirate Bay had 32 seeders per leechers and EZTV had 46 seeders per leecher, which does not even come close to the peak ratios of the private communities. In fact, in The Pirate Bay, as much as 47% of our observations were ratios below 1.

[2]This is again the overall average of the per-swarm averages.

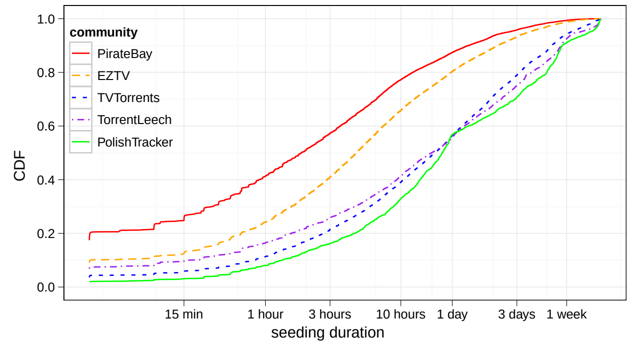

Fig. 5. The CDF of the seeding duration per community.

D. Seeding duration

As the BitTorrent protocol offers no direct incentive for seeding, communities have to rely either on altruism or on some additional incentive mechanism; the reciprocal capacity provided by leechers themselves with BitTorrent's tit-for-tat protocol works reasonably well during flashcrowds, but is insufficient for sustainable performance. The private communities in our selection therefore use credits or ratio enforcement to force their members to seed.

Fig. 5 and Table II show the CDF and several statistics of the seeding duration per community. Again, note the log-scale on the horizontal axis of the CDF. In public community The Pirate Bay, 20% of the peers do not seed at all, 44% of the peers seed for less than one hour, and only 13% of the peers seed more 1 day. EZTV has slightly higher seeding durations, with 20% of the peers seeding more than 1 day. These measured seeding durations are significantly longer than those that we measured during our experiments in 2003–2004 [10], where 83% of 53,883 measured peers were seeding for less than one hour. In the measurements of Guo *et al.* [5] in 2005, only 8% of the peers were seeding longer than 1 day.

However, the currently measured seeding durations in the public communities are still significantly lower than those in the private communities, where more than 43% of the peers are seeding longer than 1 day and even 6–9% of the peers are seeding longer than 1 week. Most extreme is PolishTracker, where only 2% of the peers do not seed at all and the majority of the peers seed for at least 20 hours.

The difference in seeding duration between the three private communities is very small, which is interesting since their policies enforce quite different minimum seeding times and ratios (see Table I). Apparently, it is most of all important that there is a ratio enforcement mechanism in place; the precise rules matter less. Consequently, the differences in download speeds observed in Section IV-A have to be due to different numbers of seeders, and/or different upload/download capacities. The seeder/leecher ratio results partly confirm this.

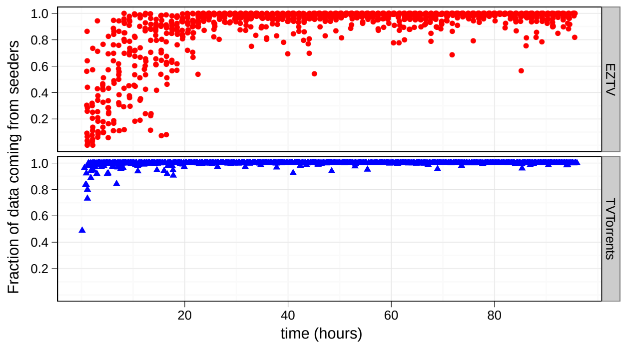

Fig. 6. The fraction of data coming from seeders in EZTV (top) and TVTorrents (bottom), over time (since swarm birth).

E. Fraction of data supplied by seeders

In BitTorrent, leechers and seeders have different upload policies [3]. Leechers prefer to upload to peers that reciprocate via the tit-for-tat mechanism, while seeders partly upload to the fastest downloaders and partly perform a round-robin selection over all interested peers. A high fraction of data supplied by seeders therefore indicates a very low contribution of tit-for-tat to the download.

Fig. 6 shows the fraction of data supplied by seeders (since swarm birth) for both EZTV (representing the public communities) and TVTorrents (representing the private communities). The results for TVTorrents show that after about 2 hours, virtually all of the data comes from seeders. Apparently, tit-for-tat is almost irrelevant in such communities. This is not so surprising, given the high seeder/leecher ratios and the high seeding durations demonstrated in Sections IV-C and IV-D. Hence, private communities are in essence more similar to systems based on direct FTP transfers than to swarming systems where downloaders also upload. This is a very important observation, since a lot of research into BitTorrent focuses on the tit-for-tat mechanism and its direct reciprocity. Many subtle optimizations and variations on this protocol are suggested (e.g., [8], [14]), but apparently such optimizations will have very limited influence when community policies such as ratio enforcement dominate users' behavior.

V. Related work

Important early measurement studies on P2P networks are of Saroiu *et al.* [12], who measure and analyze Gnutella and Napster, and of Gummadi *et al.* [4], who focus on Kazaa. Well-known early measurement studies of the BitTorrent protocol are by Izal *et al.* [7] on the evolution of a torrent; by Pouwelse *et al.* [10] on availability, integrity, flashcrowds, and performance; and by Guo *et al.* [5] on torrent popularity, torrent life-span, and multi-torrent participation. The work of Andrade *et al.* [1] presents measurement results of three communities, focusing on file popularity, supply, and demand. They find that torrent popularity distributions are non-heavy-tailed, that a small set of users contributes most of the resources, and that users that provide more resources are also those that demand more from it. More recent work is presented by Stutzbach *et al.* [13] on churn, Rasti *et al.* [11] on performance, and Mol *et al.* [9] on firewalls.

VI. Conclusions

In this paper, we have presented extensive measurements of over half a million peers in two public and three private BitTorrent communities. Our most important findings are that: (1) the download speeds in private communities are 3–5 times higher than in public communities; (2) the observed average download speeds are at least 4 times as high as those observed in 2003–2004; (3) around 47–48% of the peers in public communities are unconnectable, whereas in private communities this is only 20–34%; (4) the seeder/leecher ratios in private communities are at least 10 times as large as those in public communities; (5) peers seed for a significantly longer duration in private communities, with more than 43% of the peers seeding longer than 1 day; (6) in private communities, almost all data is supplied by seeders, therefore rendering the contribution and importance of BitTorrent's tit-for-tat mechanism virtually irrelevant.

References

[1] N. Andrade, M. Mowbray, A. Lima, G. Wagner, and M. Ripeanu. Influences on cooperation in bittorrent communities. In *P2PECON'05*, pages 111–115, New York, NY, USA, 2005. ACM.

[2] N. Andrade, E. Santosneto, F. Brasileiro, and M. Ripeanu. Resource demand and supply in bittorrent content-sharing communities. *Computer Networks*, November 2008.

[3] B. Cohen. Incentives Build Robustness in BitTorrent. In *Workshop on Economics of Peer-to-Peer Systems*, Berkeley, USA, May 2003. http://bittorrent.com.

[4] P. K. Gummadi, R. J. Dunn, S. Saroiu, S. D. Gribble, H. M. Levy, and J. Zahorjan. Measurement, modeling, and analysis of a peer-to-peer file-sharing workload. In *SOSP*, pages 314–329, 2003.

[5] L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, and X. Zhang. Measurements, analysis, and modeling of bittorrent-like systems. In *IMC'05*, pages 4–4, Berkeley, CA, USA, 2005. USENIX Association.

[6] A. Iosup, P. Garbacki, J. Pouwelse, and D. Epema. Correlating topology and path characteristics of overlay networks and the internet. In *CCGRID '06*, page 10, Washington, DC, USA, 2006. IEEE Computer Society.

[7] 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 *PAM*, pages 1–11, 2004.

[8] D. Levin, K. LaCurts, N. Spring, and B. Bhattacharjee. Bittorrent is an auction: analyzing and improving bittorrent's incentives. In *SIGCOMM'08*, pages 243–254, 2008.

[9] J. Mol, J. Pouwelse, D. Epema, and H. Sips. Free-riding, Fairness, and Firewalls in P2P File-sharing. In *IEEE P2P'08*, September 2008.

[10] J. A. Pouwelse, P. Garbacki, D. H. J. Epema, and H. J. Sips. The Bittorrent P2P File-sharing System: Measurements and Analysis. In *IPTPS'05*, volume 3640 of *LNCS*, pages 205–216. Springer-Verlag, Feb 2005.

[11] A. H. Rasti and R. Rejaie. Understanding peer-level performance in bittorrent: A measurement study. In *ICCCN*, pages 109–114. IEEE, 2007.

[12] S. Saroiu, P. K. Gummadi, and S. D. Gribble. A measurement study of peer-to-peer file sharing systems. In *MMCN '02*, 2002.

[13] D. Stutzbach and R. Rejaie. Understanding churn in peer-to-peer networks. In *IMC'06*, pages 189–202, New York, NY, USA, 2006. ACM.

[14] Y. Tian, D. Wu, and K.-W. Ng. Modeling, analysis and improvement for bittorrent-like file sharing networks. In *INFOCOM'06*, 2006.

[15] S. Xie, G. Y. Keung, and B. Li. A measurement of a large-scale peer-to-peer live video streaming system. In *ICPP'07 Workshops*, page 57, 2007.