

User behaviors in private BitTorrent communities

Jia, AL; Chen, X; Chu, X; Pouwelse, JA; Epema, DHJ

DOI

[10.1016/j.bjp.2013.12.010](https://doi.org/10.1016/j.bjp.2013.12.010)

Publication date

2014

Document Version

Accepted author manuscript

Published in

Computer Networks

Citation (APA)

Jia, AL., Chen, X., Chu, X., Pouwelse, JA., & Epema, DHJ. (2014). User behaviors in private BitTorrent communities. *Computer Networks*, 60, 34-45. <https://doi.org/10.1016/j.bjp.2013.12.010>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

User Behaviors in Private BitTorrent Communities

Adele L. Jia^a, Xiaowei Chen^b, Xiaowen Chu^b, Johan A. Pouwelse^a, Dick H.J. Epema^a

^a*Parallel and Distributed Systems, Department of Software and Computer Technology, Delft University of Technology, the Netherlands.*

^b*Department of Computer Science, Hong Kong Baptist University, Hong Kong.*

Abstract

Many private BitTorrent communities employ Sharing Ratio Enforcement (SRE) schemes to incentivize users to contribute. It has been demonstrated that users in private communities are highly dedicated and that they seed much longer than users in communities where SRE is not employed. While most previous studies focus on showing the positive effect of user dedication in achieving high download speed, in this paper we explore the user behaviors in private communities, we argue the reasons for these behaviors, and we demonstrate both the positive and the negative effects of these behaviors. We show that under SRE, users seed for excessively long times to maintain required sharing ratios, but that their seedings are often not very productive (in terms of low upload speed) and that their long seeding times do not necessarily lead to large upload amounts. We find that as users evolve in the community, some users become more committed, in terms of increasing ratios between their seeding and leeching times. In the mean time, some users game the system by keeping risky and low sharing ratios while leeching more often than seeding. Based on these observations, we analyze strategies that alleviate the negative effects of these user behaviors from both the user's and the community administrator's perspective.

Keywords:

Private BitTorrent community, user behavior, Sharing Ratio Enforcement, incentive policy, Peer-to-Peer.

1. Introduction

BitTorrent is a popular Peer-to-Peer (P2P) protocol for file distribution. A key to its success lies in its Tit-For-Tat (TFT) incentive policy, which works reasonably well in fostering cooperation among downloading peers¹ (also known as *leechers*). However, TFT does not provide any incentive for peers to remain

¹In this paper, we use the term *user* and *peer* alternatively to refer to the individuals in a BitTorrent community.

in the system after the download is complete in order to *seed* the entire file to others. Therefore, peers are free to engage in “Hit and Run”, the scenario under which a peer leaves immediately upon completing a download. To provide incentives for seeding, in recent years there has been a proliferation of so-called *private* BitTorrent communities. It has been demonstrated that users in private communities are highly dedicated [1, 2, 3, 4]. While most previous studies focus on showing the positive effect of user dedication in achieving high download speed, in this paper we explore the user behaviors in private communities, we argue the reasons for these behaviors, and we demonstrate both the positive and negative effects of these behaviors.

Most private communities employ a private-tracker-based method that maintains centralized accounts and records the *sharing ratio* of each peer, i.e., the ratio between its total amount of upload and download. Community administrators specify some *threshold* above which all members are required to maintain their sharing ratios. This mechanism is known as *Sharing Ratio Enforcement* (SRE). Community members whose sharing ratios drop below the threshold are warned and then banned from downloading, or even expelled from the community. In this way, it is guaranteed that each peer provides a certain level of contribution to the community.

The main motivation for implementing SRE is to close the gap between bandwidth demand and supply as observed in public communities, where there is significantly more demand than supply [3]. Thus, the basic design goal of SRE is to achieve higher system-wide download speed by encouraging users to contribute and therefore increasing the bandwidth supply. Several measurement studies have shown that SRE is very effective in incentivizing users to contribute [1, 2, 3, 4, 5]. For instance, [3] reports seeder-to-leecher ratios that are at least 9 times higher in private communities than in public ones, while download speeds are found to be 3-5 times higher. Nevertheless, why are users in private communities more dedicated? Is the dedication due to user’s internal willingness or is forced by the external environment? Besides the universal dedication, are there other types of user behavior? And besides the high download speed, are there any other positive or even negative effects of these user behaviors?

To answer these questions, in this paper we explore the user behaviors in private communities. We classify users into different groups based on their sharing ratios, their ages, their levels of consumption, and their *effort ratios*. The effort ratio of a user is defined as the ratio between his seeding and leeching time. We demonstrate the behavioral differences between users in different groups, we argue the reasons for these differences, and we show the positive and negative effects of these behaviors based on metrics including the seeding time, the upload speed, and the evolution of sharing ratio, which are highly related to user behaviors. Our previous work [5, 6, 7, 8] has dealt with analyzing the advantages and disadvantages of SRE schemes based on theoretical models, simulations, and measurements. This paper complements these works by presenting observations from real world communities with a focus on the user behavior patterns.

The main contributions of this paper are as follows:

1. We perform a measurement study of three private communities that provide user-level information including the upload amount, the download amount, the seeding time, the leeching time, and the sharing ratio of *each* individual user. Among the dozens of existing communities we have examined, these are the only ones that provide such detailed information. We use one of the three communities as an example to explore the user behaviors.
2. Based on the sharing ratio we classify users into the *rich*, the *middle class* and the *poor*. We show that, to maintain adequate sharing ratios, all users have to seed for excessively long times (compared to their downloading times), though most of the time their seedings are not very productive and their long seeding times do not necessarily lead to large upload amounts. For users who intend to increase their sharing ratios, we find that seeding for longer durations is not as effective as increasing the upload speed, which can be achieved by upgrading the internet access, or as joining swarms in their early stages to avoid situations of oversupply.
3. Based on the age we classify users into the *new* and the *old*. We find that old users are more committed to the community, in terms of higher ratios between their seeding and leeching times, while new users in general seed more productively.
4. Based on the download amount we classify users into the *big* and the *small consumers*. We find that big consumers are often at the same time big contributors, and are more active than small consumers in terms of both longer seeding times and longer leeching times.
5. Based on the effort ratio we classify users into *gamers* and *dedicators*. We find that gamers not only leech longer but also seed shorter than dedicators, and at the same time they maintain lower sharing ratios, which, however, are still high enough for them to stay in the community.
6. Based on the user behaviors we defined, we analyze strategies that alleviate the negative effects of these user behaviors from both the users' and the community administrators' perspective.

2. Methodology

In order to obtain a better understanding of private BitTorrent communities it is critical to be able to collect data on their operation. Over the years it has been proven to be a challenge to obtain detailed traces of user behavior, and due to a combination of technical constraints and privacy concerns, few prior work was able to capture both detailed user profiles, content availability, and precise information on every user download.

To support our analysis, we have examined 38 elite private communities, out of which we selected three communities, CHDBits [9], HDStar [10]), and ChinaHDTV [11], for detailed regular deep crawling of HTML pages. These specific three communities were selected as they are the only ones that provide information detailed enough for our analysis. We have obtained the following three datasets for each community:

1. **Community-level user profile (dataset 1):** in this dataset, we crawl the profile page of *each* community user and obtain the information of his upload and download amount, his seeding and leeching time, his sharing ratio at the time of snapshot, and the time he joined the community. It should be noted that the seeding time of each user recorded by the tracker is *swarm-based*, i.e., simultaneously seeding in multiple swarms counts separately. For instance, after a user has seeded in two swarms for 10 hours, $2 \times 10 = 20$ hours will be added to his seeding time. Similarly, the leeching time recorded by the tracker is also swarm-based. In later sections, when we calculate the average upload speed of a user, we calculate his per-swarm average upload speed, i.e., the total upload amount divided by the swarm-based seeding time. In this way, we get a rough estimation of a user's seeding time and upload speed. Though more accurate calculation of the seeding time and upload speed would be better, to the best of our knowledge, until now no private communities provide this information. Further, it is almost impossible to deploy a client and contact the tracker or every user individually to get this information due to the risk of being expelled from the community.
 2. **Community-level torrent profile (dataset 2):** in this dataset, we crawl the community trackers and collect information of *each* torrent, including the number of seeders and leechers, the number of finished downloads at the time of the snapshot, and the time the torrent was published.
 3. **Torrent-level user activity (dataset 3):** the tracker records a user's torrent-level action times, such as the time of joining the swarm, the time of starting seeding, etc. The precision of the recorded action time decreases with time. For example, if a user started to seed 10 hours ago, its action time will be "10 hours ago". However, if a user started to seed one month, 23 days, and 10 hours ago, its action time will only be "one month and 23 days ago".
- In order to obtain the action times with precision in hours, for each community, we examine all the torrents released within 24 hours. We follow these torrents for 7 days and record the activity of *each* user who has participated or is participating in one of them. The collected information includes each user's per-swarm upload amount, download amount, seeding time, and leeching time, as well as the time he joins and leaves the swarm.

Private communities often consist of tens of thousands of torrents and users. For instance, when we collected the data in May, 2011, CHDBits had 33,872 torrents and 29,548 registered users out of which 24,633 users maintain public profile pages that can be viewed by others. HDStar had 7,809 torrents and 14,984 registered users (out of which 13,593 users are public), and HDTV had 6,234 torrents and 92,100 registered users (out of which 91,815 users are public)². We have analyzed the measurements of the three communities in detail. As we

²Comparing to the other two communities, HDTV had a large user base but a small number of torrents. We conjecture that the possible reason could be that HDTV does not delete the

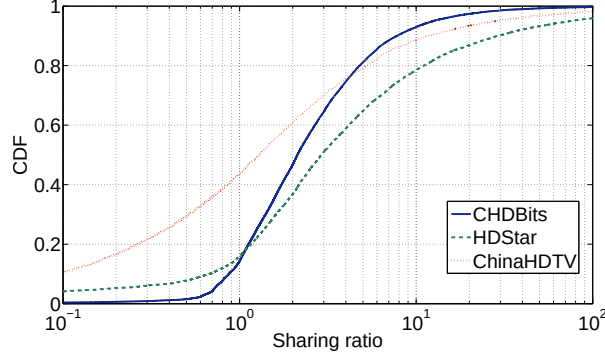


Figure 1: The CDF of the user sharing ratio in CHDBits, in HDStar, and in ChinaHDTV. The horizontal axis is in log scale.

show later in Section 3, they demonstrate similar performance. For simplicity of presentation, we only demonstrate the results of CHDBits in Sections 4 and 6.

The detailed user behavior information allows us to explore the user behavior in private communities. User behavior directly decides a user’s basic movements including (1) when to join the community, (2) how much to consume, and (3) how much to contribute, from which it further decides (4) a user’s status in the community, in terms of the sharing ratio he achieves. Based on these four metrics we classify users into different groups, we demonstrate their behavioral differences, we argue the reasons for these differences, and we show the positive and negative effects of these behaviors. We start from the sharing ratio, since without a proper one a user cannot even stay in the community, which implies its fundamental importance.

3. The rich and the poor: positive and negative effects of SRE

In this section, we divide users into different groups based on their sharing ratios. We analyze the reasons for some users to achieve low sharing ratios, from which we demonstrate the positive and the negative effects of SRE.

3.1. A general view

We first show in Fig. 1 the CDF of the user sharing ratio in CHDBits, in HDStar, and in ChinaHDTV, respectively (dataset 1). We see that most users in these three communities achieve sharing ratios larger than the SRE threshold, i.e., 0.7. Take users in CHDBits for example, around 15% users have sharing ratios less than 1 (defined as the *poor*), while around 18% users have sharing

profile pages of users who have left the community permanently. Unfortunately, we don’t have the information of how many such users there are.

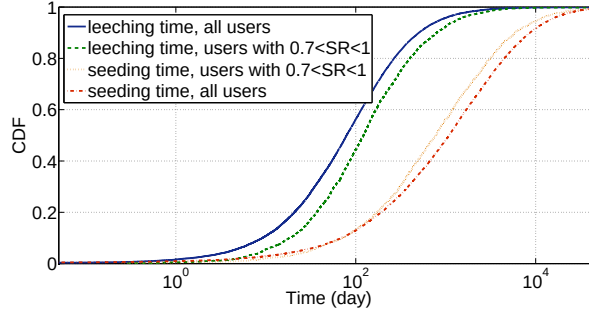


Figure 2: The CDF of the seeding and the leeching time in CHDBits. The horizontal axis is in log scale.

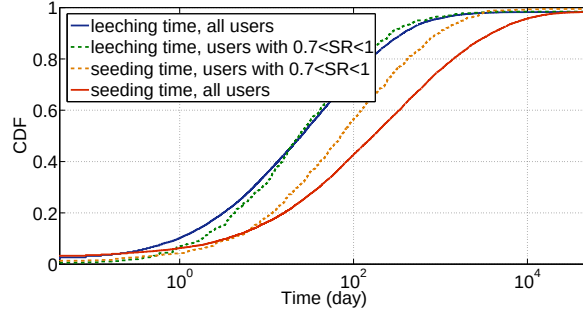


Figure 3: The CDF of the seeding and the leeching time in HDStar. The horizontal axis is in log scale.

ratios larger than 5 (defined as the *rich*). The rest that have sharing ratios between 1 and 5 are defined as the *middle class*. The behavior of accumulating a large sharing ratio may be triggered by various motivations, such as altruism, a desire to be part of the rich elite of the community, or a habit of saving sharing ratio for the future. The rich peers have little worry about staying in the community, since their sharing ratios are far beyond the SRE threshold. On the other hand, poor peers are at the risk of being expelled from the community. As a consequence, they need to be concerned a lot about their decisions: they may download new contents they really desire, but this might reduce their sharing ratios to a more risky level.

One may argue that the poor peers are free-riders, who intend to keep low and risky sharing ratios that are just enough to stay in the community. However, the highly restricted membership in private communities, especially in CHDBits and many other private communities where new members can only join by a limited number of invitations, makes it very difficult to get a new membership. Hence, we conjecture that not all poor peers are strategic and psychologically strong enough to face being expelled from the community due to insufficient sharing ratios. Interestingly, as we will show in the following sections, the

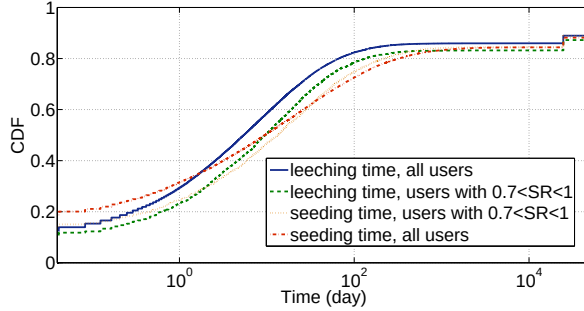


Figure 4: The CDF of the seeding and the leeching time in ChinaHDTV. The horizontal axis is in log scale.

poverty is partially induced by the fact that the poor peers are not strategic enough.

3.2. Long seeding time, even for the poor

Many previous studies have shown that under SRE, users seed for long durations [1, 2, 3, 4]. They consider this as a positive effect of SRE since long seeding durations lead to high download speeds. However, in this section we argue that the long seeding durations can also be seen as a negative effect, especially for poor peers.

Figs. 2, 3, and 4 show the CDFs of the seeding time and the leeching time in CHDBits, in HDStar, and in ChinaHDTV, respectively (dataset 1). Consistent with the theoretical results of our previous work [7, 8], in general the seeding time is much longer than the leeching time in all these three communities. Take CHDBits for example, the median leeching time is 70 days while the median seeding time is 1,100 days. Remember that the seeding and the leeching time of users are swarm-based, leading to very high values.

Intuitively, longer seeding times than leeching times for rich peers are to be expected, since rich peers are saving sharing ratios by seeding. However, we observe from Figs. 2, 3, and 4 that, even poor users seed much longer than they leech. While intuitively poor peers should be the ones that are not “hard-working” enough, why do some of them seed for long durations but still have low sharing ratios? In the following section, we explore the possible reasons.

As shown in Figs. 1 to 4, users in these three communities behave similarly. For simplicity of presentation, from now on we only show the results for CHDBits, without explicitly stating so. For the results shown in later sections, the datasets of the other two communities demonstrate similar performance as that of CHDBits [12].

3.3. Possible reasons?

One may argue that the long seeding times of poor peers are due to the fact that even though they contribute more, they also consume more. Hence, they

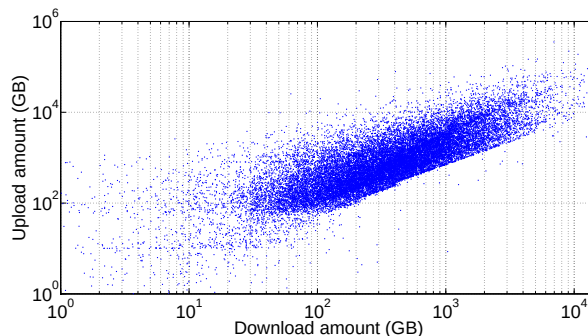


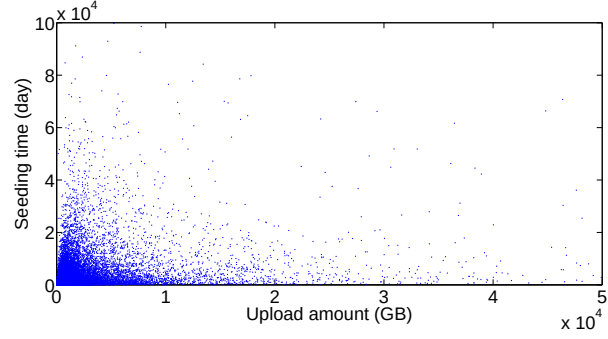
Figure 5: The upload amount versus the download amount (with Spearman’s rank correlation coefficient equal to 0.8110). The horizontal axis is in log scale.

seed for long durations but they still have low sharing ratios. This argument is partially true. Andrade *et al.* [13] have shown and we also observe from our measurement (Fig. 5, dataset 1) that the individual upload amount (contribution) increases with the corresponding download amount (consumption), with the Spearman’s rank correlation coefficient equal to 0.8110. Spearman’s rank correlation coefficient assesses how well the relationship between two variables can be described using a monotonic function [14]. However, this doesn’t necessarily mean that heavy contributions induce long seeding times, nor does it mean that long seeding times lead to heavy contributions.

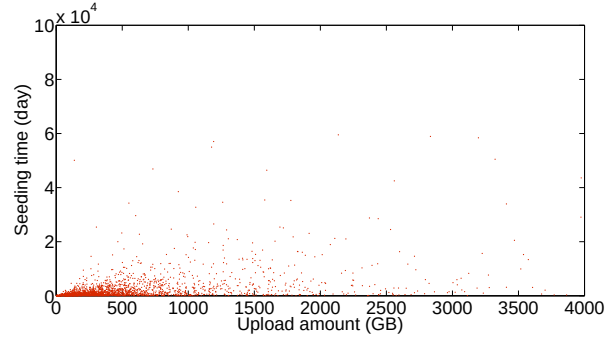
Quite counter-intuitively, as shown in Fig. 6(a), a peer’s upload amount has little relation to its seeding time: many peers seed for long durations but only have uploaded relatively small amounts of data, while other peers seed for relatively short durations but have successfully achieved large upload amounts. The same argument is also applicable to poor peers (Fig. 6(b)). This interesting phenomenon implies that for poor peers who intend to increase their upload amount to become rich, seeding for longer durations may not be an effective method, even if intuitively it seems so.

Though there is no strict relationship between a peer’s seeding time and its upload amount, we do observe that a peer’s seeding time is related to its average upload speed, regardless of its upload amount. As shown in Fig. 7, most of the long seeding durations happen to the peers with relatively small upload speeds, and for peers who have high upload speeds, the seeding times are normally short.

The most intuitive reason for a low upload speed is a limited internet access. However, we argue that this is not the only reason. From dataset 2, at the time when we crawled the site, CHDBits had 33,041 active swarms (with at least one leecher or one seeder), among which 26,402 swarms (79.9%) had no leechers at all. As shown in Fig. 8(a), 40% of the swarms with no leechers still have at least 5 seeders, and 5% of these swarms even have more than 20 seeders. For swarms with at least 1 leecher, the seeder-to-leecher ratio (SLR) is quite high: as shown in Fig. 8(b), 50% of these swarms have as SLR larger



(a) for all users



(b) for users with $0.7 \leq SR \leq 1$

Figure 6: The seeding time versus the upload amount.

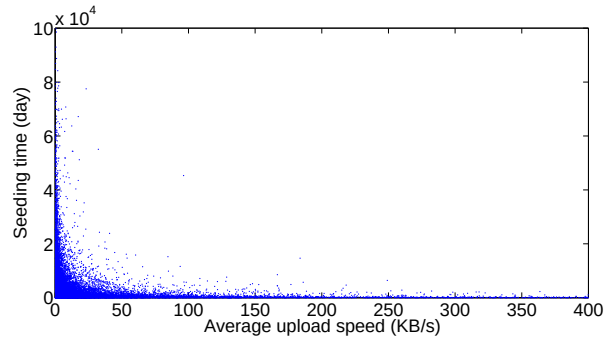


Figure 7: The seeding time versus the upload speed (with Spearman's rank correlation coefficient equal to -0.6318).

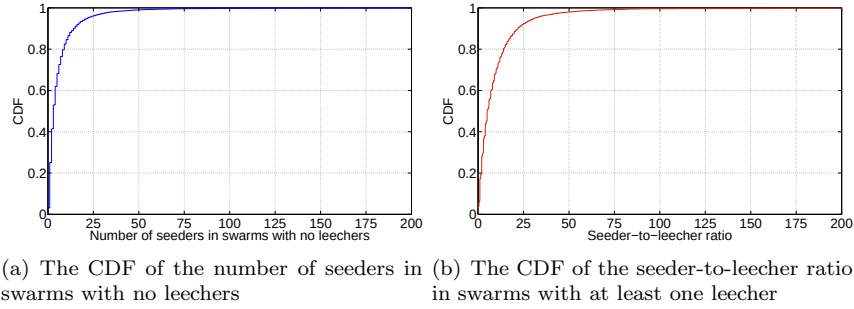


Figure 8: The oversupply in swarms

than 6, and 5% of these swarms even have as SLR larger than 30. We see clearly that a majority of the swarms in CHDBits are heavily oversupplied. In such swarms, seeders are not able to perform any actual uploads due to the insufficient demand and unsatisfied supply. We term this situation *unproductive seeding*. As a consequence, users have to seed for excessively long durations to achieve the sharing ratio required by SRE.

While a low upload speed mainly leads to a long seeding time, in the next section we show its influence on a user’s status. We analyze the reasons for the poor being poor and discusses strategies for users to become rich efficiently.

3.4. Why the poor are poor and how to become rich?

As the sharing ratio is defined as the ratio between a peer’s upload and download amount, two possible reasons for a peer being poor are that it has downloaded too much or has uploaded not enough. The download amount depends on a user’s interests in contents. We do not suggest users to download less so as to become rich, since the fundamental user experience that should be guaranteed by communities is that users should not need to limit their download needs. Following this argument, in this section we focus on the user upload activity and analyze why some users have uploaded not enough (hence, are poor) and how they can improve it (to become rich).

3.4.1. Community level

In Section 3.3 we have shown that the seeding time has little influence on the upload amount but the upload speed does. The upload speed further influences whether a user is rich or poor. As shown in Fig. 9 (dataset 1), in general rich peers ($SR \geq 5$) have much higher upload speeds than poor peers ($SR \leq 1$). For example, 80% of the poor peers upload at a speed less than 20 KB/s, while at least 40% rich peers can upload at a speed larger than 50 KB/s. Together with the result in Section 3.3, we conclude that instead of seeding for longer durations, peers who intend to become rich should seed with higher upload speeds. And to seed with a higher upload speed, a user could upgrade its internet access or choose a swarm that is less oversupplied.

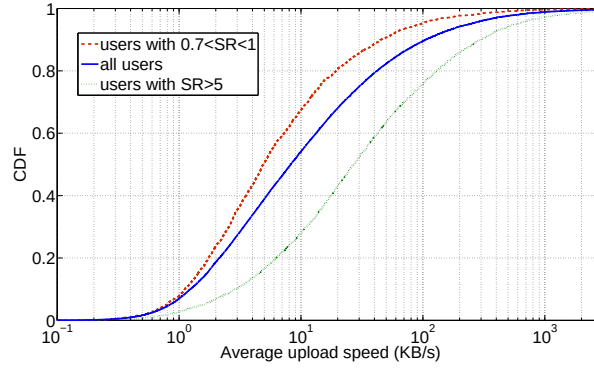


Figure 9: The CDF of the average upload speed. The horizontal axis is in log scale.

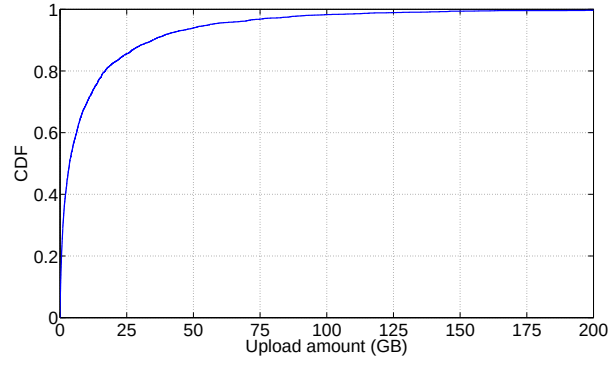


Figure 10: The CDF of the upload amount in one swarm.

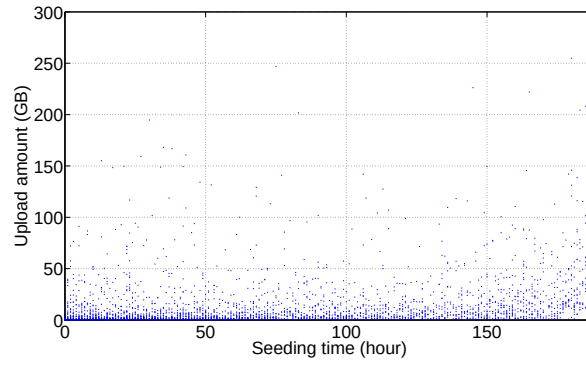


Figure 11: The upload amount versus the seeding time in one swarm.

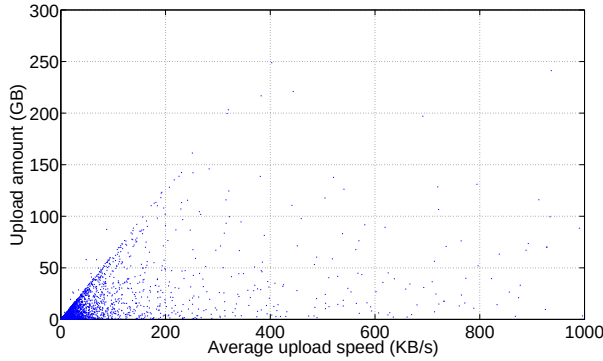


Figure 12: The upload amount versus the upload speed in one swarm (with Spearman’s rank correlation coefficient equal to 0.7876).

One may argue that the above analysis is based on community-level activities, which only provide a macroscopic view that is not enough to show the underlying details. To explore this, in the following subsection we focus on a single swarm and demonstrate the torrent-level user performance, and we discuss possible strategies for users to become rich.

3.4.2. Torrent level

Among all the CHDBits swarms in dataset 3, we choose the one with the largest number of participants as the example to show the torrent level user behaviors. This file has a size of 11.6 GB. In total, 3,776 users are included.

Different individual upload amount in one swarm: Fig. 10 shows the CDF of the user upload amount in a single swarm (dataset 3), from which we observe that a small fraction of users have uploaded considerably more than the others. For example, 60% of the users have uploaded less than 10 GB, which is less than the amount they have downloaded (11.7 GB). On the other hand, 5% of the users have uploaded more than 50 GB. Of course, the users who managed to upload more will become richer. While these users have participated in the very same swarm, why did some manage to gain a lot while others didn’t?

Possible reasons and how to gain more: One intuitive reason for a small upload amount is a short seeding time. However, similar to the analysis in Section 3.3, again we find the counter-intuitive result that in one swarm a peer’s upload amount is not related to its seeding time (Fig. 11). On the other hand, it is related to its upload speed. As shown in Fig. 12, most of the small upload amounts happen to the peers with relatively low upload speeds, and peers with high upload speeds normally have uploaded a large amount.

When we organize the peers according to the time they start to seed, we find another interesting phenomenon: peers that start to seed earlier normally have uploaded more (Fig. 13). The same phenomenon has also been observed by Kash *et al.* in [15]. One may argue that the peers who start to seed earlier can seed for longer durations, hence they upload more. However, in Fig. 11 we

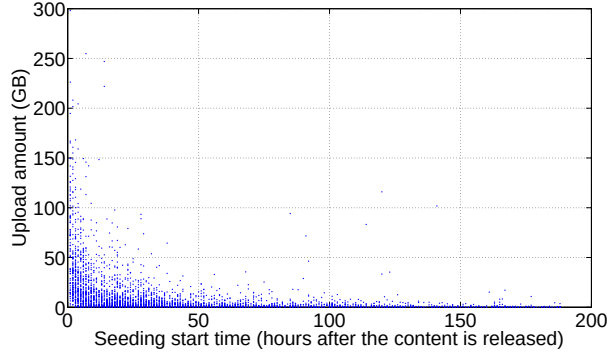


Figure 13: The upload amount versus the time of starting seeding in one swarm (with Spearman’s rank correlation coefficient equal to -0.6491).

already show that the upload amount is not related to the seeding time. Then why do peers that start to seed earlier upload more?

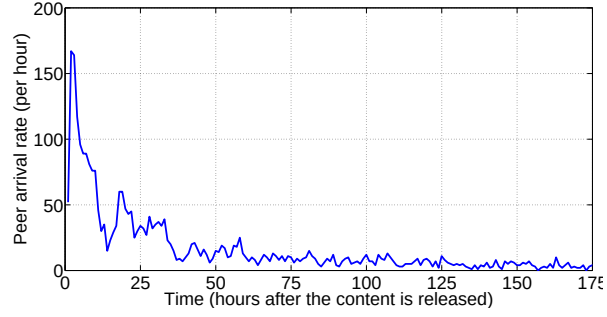
As shown in Fig. 14(a), after the burst at the first two hours since the file was published, the peer arrival rate decreases dramatically. On the other hand, the number of seeders increases quickly at the first 60 hours, then decreases with a much smaller rate (Fig. 14(b)). In general, the number of leechers is negligible compared to the number of seeders. As a consequence, peers who join late have to compete with a large number of seeders for uploading, which leads to a low upload speed, and hence a small upload amount. The burst of new peers in the early state of the swarm, or *flashcrowd*, is a common phenomenon that has been observed in many BitTorrent swarms [16, 17]. Therefore, peers who intend to become richer should join the swarm in its early stage, when it is still not extremely oversupplied.

4. The old and the new: how users evolve

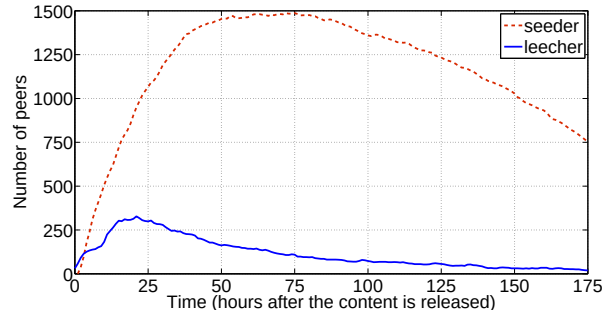
The first behavior of any user is to register as a member and join the community. In this section, we explore the behavioral differences between users of different ages based on community-level user profiles (dataset 1) we obtained.

4.1. A general view

After joining the community, users gradually build up a history of uploads and downloads, in terms of the amount they consume and contribute, and the time they spend in leeching and in seeding. Fig. 15 shows the scatter plot of the upload and the download amount of each user, with the users ranked in the reverse order of their ages. We see that, in general, the upload and the download amount of the first 20,000 users are stable with their ages, while the remaining ones demonstrate a decreasing trend when the age decreases. We observe a similar correlation between the seeding and the leeching time of users and their ages, as shown in Fig. 16. With these clear behavioral differences we



(a) The peer arrival rate (per hour)



(b) The evolution of the numbers of leechers and seeders

Figure 14: The performance in one swarm.

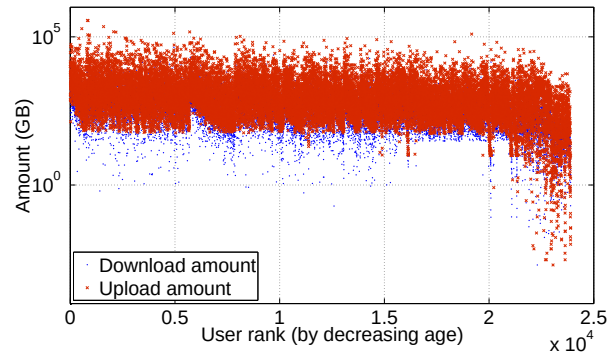


Figure 15: The upload and the download amount versus the age. The vertical axis is in log scale.

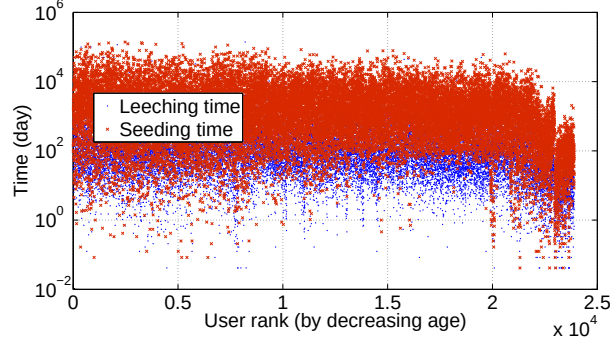


Figure 16: The seeding and the leeching time versus the age. The vertical axis is in log scale.

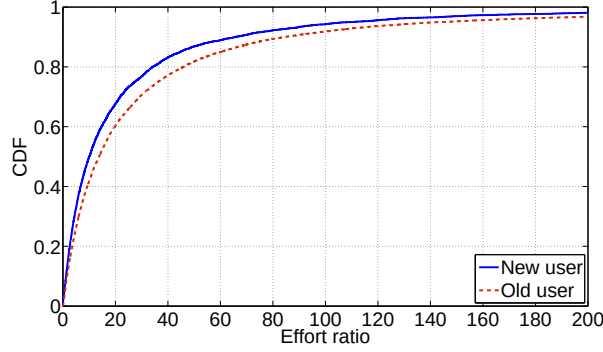


Figure 17: The CDF of the effort ratio of old and new users.

divide users into two groups, the first 20,000 users as *the old* and the other ones as *the new*. Next, we explore more behavioral differences between these two groups of users.

4.2. Level of commitment: on the way to become more committed

The most straightforward measure for a user's commitment to the community is its upload amount, which is exclusively decided by the seeding time and the average upload speed. As discussed in Section 3.4, the average upload speed is not a simple reflection of user's bandwidth, but an outcome of the swarm status like the number of seeders and leechers in the swarm. As it is difficult for users to control their average upload speeds, we use only the seeding time to reflect their level of dedication. Further, to avoid the cumulative effect of age, i.e., that older users have stayed in the community longer and therefore have higher opportunities to seed and leech longer, we use the ratio between the seeding and the leeching time, which is defined as the *effort ratio*, as the metric to measure user dedication to the community.

In Fig. 17 we show the CDF of the effort ratio of old and new users, respectively. We see that most users (both the old and the new) achieve very high

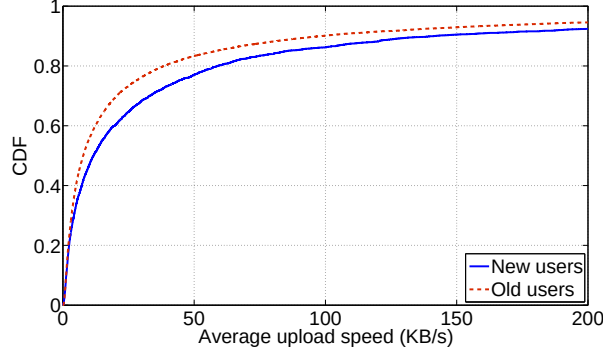


Figure 18: The CDF of the average upload speed of old and new users.

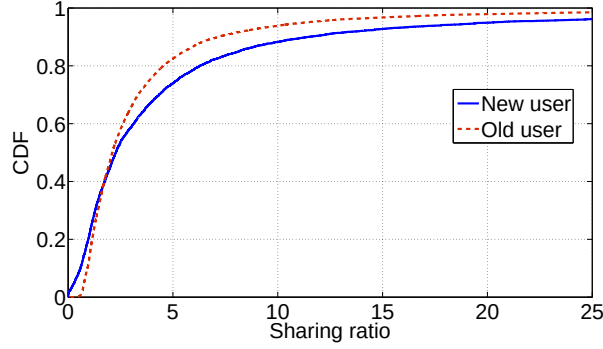


Figure 19: The CDF of the sharing ratio of old and new users.

effort ratios, indicating that with SRE, users in private communities are highly committed (or forced to be). More interestingly, in general, new users achieve lower effort ratios than old users, indicating that as users evolve in the community, they become more committed, which implies a deeper (psychological) effect of SRE on the old users.

4.3. Average upload speed: the new have not yet suffered

Under SRE, users will be expelled if they cannot upload enough to meet the SRE requirement. Intuitively, users with relatively low bandwidth will gradually become extinct, leaving only the ones with high bandwidth to survive in the community. Nevertheless, Fig. 18 shows the CDF of the average upload speed achieved by old and new users, respectively, from which we see that in general new users achieve higher upload speeds. This counter-intuitive observation confirms our analysis in Section. 3.4 that a user's average upload speed highly depends on the status of swarms it has participated in. When a user with a high bandwidth spends long time in unproductive seeding, it can achieve a very low upload speed. We have shown in Fig. 17 that new users are less

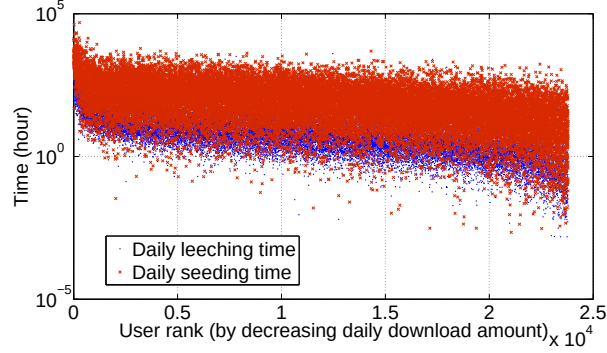


Figure 20: The daily seeding and leeching time versus the daily download amount. The vertical axis is in log scale.

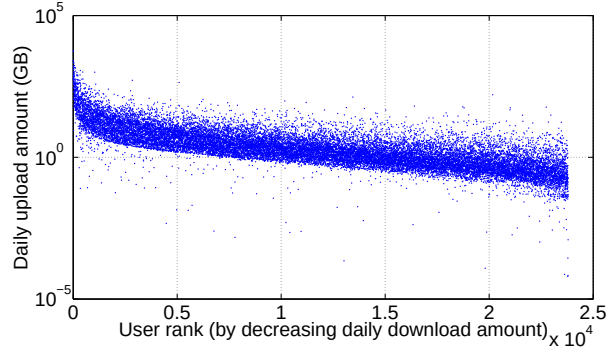


Figure 21: The daily upload amount versus the daily download amount. The vertical axis is in log scale.

dedicated to the community. We conjecture that due to this reason, they avoid long unproductive seeding time and achieve higher upload speeds.

4.4. Sharing ratio: the spreading new and the conservative old

As discussed in Section 3.4, there are both internal and external reasons for users being poor or rich. In Fig. 19 we show the CDF of the sharing ratio achieved by new and old users, respectively. We see that the line for new users is more skewed, indicating that new user achieve a larger range of sharing ratios than old users. The lower end is due to the fact that new users are often given start-up time to increase their sharing ratios, while the higher end is normally due to their small download amounts. On the other hand, old users behave conservatively, without many risky sharing ratios below the SRE threshold or excessive sharing ratios requiring huge contribution and little consumption.

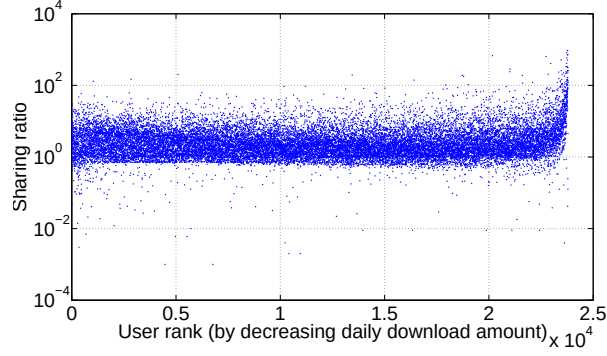


Figure 22: The sharing ratio versus the daily download amount. The vertical axis is in log scale.

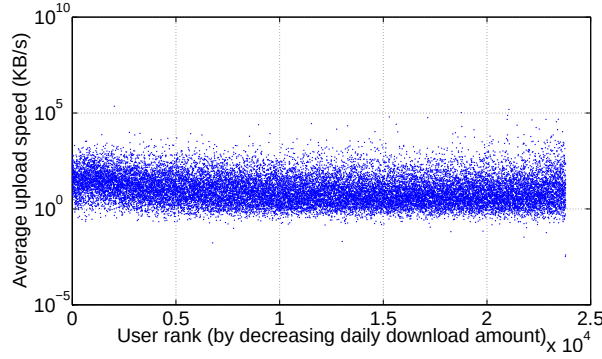


Figure 23: The average upload speed versus the daily download amount. The vertical axis is in log scale.

5. The big and the small consumer: active users active in all

How much to consume is a fundamental user behavior in private communities, since the main goal for users to join is to download the contents they are interested in. In this section, we explore the behavioral differences between users with different download amounts based on community-level user profiles (dataset 1) we obtained.. As discussed in Section 4, a user's age has a cumulative effect on his download and upload amounts, and on his seeding and leeching times. This implies that as users evolve in the system, these four metrics gradually increase. To avoid this cumulative effect, we consider the daily values for these metrics.

Figs. 20 and 21 show the daily seeding time, the daily leeching time, and the daily upload amount of each user, with users ranked according to the decreasing order of their daily download amounts. We see a clear decreasing trend of these three metrics when the daily download amount is decreased, indicating that users with larger consumptions normally contribute more, in terms of both the

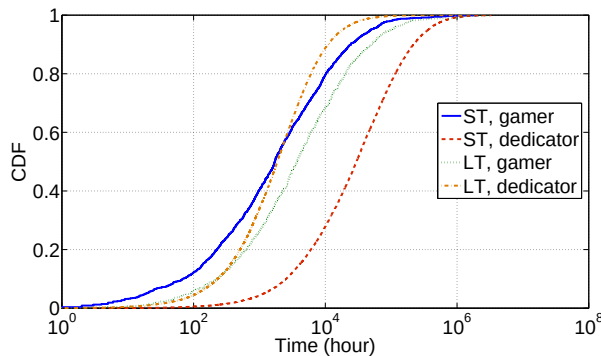


Figure 24: The CDFs of the seeding and the leeching time of gamers and dedicators. The horizontal axis is in log scale.

time and the amount they contribute. This also implies that active users are active in both downloading and uploading.

In Figs. 22 and 23 we show the sharing ratio and the average upload speed for each user, with users again ranked according to the decreasing order of their daily download amounts. This time we observe no clear correlation between these metrics, except that some users at the right end achieve extremely large sharing ratios. We believe this is due to their extremely small daily download amount.

6. The gamer and the dedicator: Gemini of the private community

How much to contribute is an important choice users make in private communities. Based on the level of contribution, a user can be a *gamer* who games the system, explores the potential benefits, and avoids providing much contribution, or a *dedicator* who dedicates himself to the system and provides high contribution—two opposite user behaviors evolved in the same private community, just like the Gemini.

The upload amount is often used to measure a user’s contribution level. Nevertheless, while users can decide the time they contribute, they do not have full control of the upload speed. Therefore, as in Section 4.2, we again use the effort ratio as the metric for deciding a user’s contribution level, and we define users with effort ratio less than one as gamers and the rest as dedicators. Next, we explore more behavioral differences between these two groups of users based on the community-level user profiles (dataset 1) we obtained.

6.1. Seeding and leeching time: which decides a gamer?

The reasons for a low effort ratio are a short seeding time and/or a long leeching time. In Fig. 24 we show the CDFs of the seeding and the leeching time of gamers and dedicators, respectively. We see that, in general, gamers leech longer while they also seed shorter than dedicators, indicating that gamers

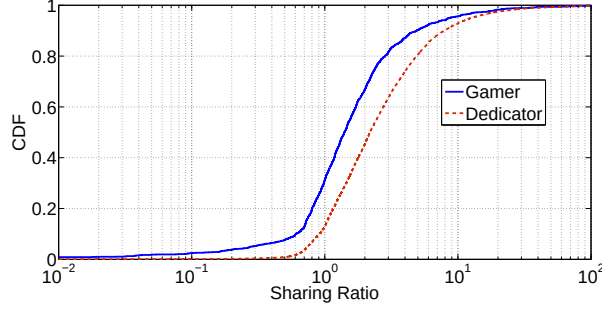


Figure 25: The CDFs of the sharing ratio of gamers and dedicators. The horizontal axis is in log scale.

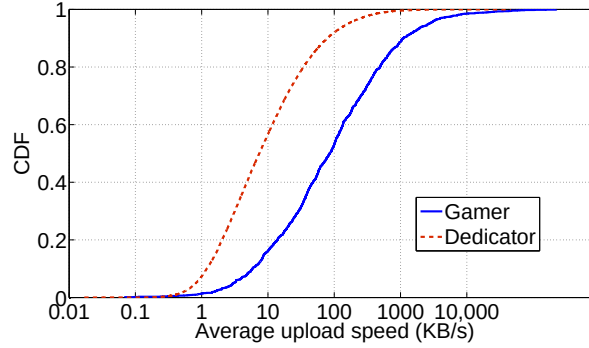


Figure 26: The CDFs of the average upload speed of gamers and dedicators. The horizontal axis is in log scale.

not only put less effort in contributing to the community but also potentially consume more.

6.2. Sharing ratio: gamers seize the day

In Fig. 25 we show the CDF of the sharing ratio achieved by gamers and dedicators, respectively. We see that in general, gamers achieve much lower sharing ratios. As the sharing ratio in a private community serves as virtual credit that can be spent in future downloads, the difference in sharing ratio implies that gamers do not hoard sharing ratios for future downloads as much as dedicators do. In other words, they are more the seize-the-day type: they keep sharing ratios that are just enough for them to stay in the community, so that they could explore the benefit of downloading (reflected by the long leeching time) and in the mean time provide little contribution (reflected by the short seeding time).

6.3. Average upload speed: gamers know the way

Though gamers achieve lower sharing ratios than dedicators, they still need to meet the SRE threshold to stay in the community. With the short seeding

time and the long leeching time, gamers have to increase their upload speed efficiently so as to maintain their sharing ratios above the SRE threshold. Consistent with our intuition, they do achieve higher upload speed than dedicators. As shown in Fig. 26, apparently, gamers can achieve upload speeds an order of magnitude higher than dedicators. As discussed in Section 3.4, achieving high upload speed often requires users to be strategic such as joining swarms early, which implies that gamers not only intend to, but also know the way, to game the system.

7. Discussion

Though altruistic users always exist, we conjecture that most users in private communities are *selfish*. Their initial goal in a community is to download all the contents they are interested in. To achieve this, they try to maintain the required sharing ratio while not limiting their download needs. The strategies they apply mainly optimize their own benefit, without considering the *social welfare*, i.e., the performance of other users. For example, users may seed all the files they have downloaded to increase the opportunity of performing some actual uploading during seeding. However, this directly increases the bandwidth supply and makes the upload competition even more severe [18]. As we discussed in Section 3.4, joining swarms earlier helps users gain sharing ratios more efficiently. However, if a majority of users strategically join a swarm immediately after a new content is published, then 1) many users will download something they don't want, only for gaining sharing ratios; 2) the download speed in the early stage of a swarm will be very low, because a large number of strategic users joining simultaneously makes the swarm heavily *flash-crowded*³; and 3) it will be more difficult to perform any actual uploads after the early stage, since only a few non-strategic users will join the swarm during that period.

Private community administrators that intend to adopt strategies, or remedies, to alleviate the side-effects of SRE, should take the potential strategic user behaviors into account. For example, some private communities try to further incentivize contribution beyond SRE by giving rich peers priority to access newly published contents [19]. However, as discussed previously, joining early in a new swarm will help the users, especially the poor users, gain sharing ratios more efficiently. By giving priority to rich peers, administrators are basically taking the opportunities away from the poor peers for gaining sharing ratios. Unless the administrators intend to let the rich be richer and the poor be poorer (which will lead to a more intense competition and a potential deterioration of performance as discussed previously), we suggest administrators to remove these restrictions.

Another example of existing remedies for SRE adopted by private communities would be *free-leech* and *seeding-bonus*. Some communities [20, 9, 11, 10]

³We refer a swarm to be flash-crowded when it has a sudden increase in the number of leechers.

temporarily adopt free-leech and/or seeding-bonus for certain swarms, which means that a user can download the file for free and/or get extra bonus for seeding. In free-leech periods, users are attracted to those swarms because of the low price for downloading. In this way, the bandwidth demand is increased and the oversupply in the system is alleviated. Meanwhile, when free-leech is applied to a relatively old swarm, the benefit of joining early is also reduced. The same argument is also applicable to seeding-bonus. In seeding-bonus periods, peers are attracted to the swarms to seed. Hence, when seeding-bonus is applied to old swarms, the file availability is improved. However, administrators should be careful and not adopt free-leech or seeding-bonus for a long time, otherwise strategic users might wait and not download anything until the files are for free, or only seed in swarms with seeding bonus.

In our previous work [7], we propose a self-organizing strategy named *SRE with supply-based price* that prevents this potential manipulation of strategic users. Instead of manually adopting free-leech (i.e., zero price), this strategy inversely relates the price for downloading one unit of data to the seeder-to-leecher ratio in the swarm. With a larger seeder-to-leecher ratio, i.e., an increasing supply, SRE with supply-based price automatically decreases the price. Once the supply goes tight again, it will automatically increase the price. In this way, the demand and supply are automatically balanced and reasonable downloading and seeding times are achieved.

8. Related work

To date, only few works have analyzed private communities. Zhang *et al.* [4] investigate hundreds of private trackers and depict a broad and clear picture of the private community landscape. Chen *et al.* [1] compare system behaviors among 13 private trackers and 2 public trackers, and they show their differences regarding user viscosity, single torrent evolution, user behaviors, and content distribution. Cuevas *et al.* [21] particularly demonstrate content publishing patterns in both public and private communities. Liu *et al.* [2] also perform measurement studies and further develop a model to show that SRE indeed provides effective incentives, but is vulnerable to collusion.

While these studies all focus on demonstrating the high seeding level achieved by private communities, there have been a few preliminary works that show the adverse effects. Andrade *et al.* [13] focus on the dynamics of resource demand and supply, and they show that users typically try to increase their contribution levels by seeding for longer and not by providing more bandwidth to the system. However, our paper shows that providing limited bandwidth is not the will of users, but it is a consequence of the oversupply in private communities. Chen *et al.* [22] also notice the oversupply problem and provide a model to identify the optimal stable SLR range. However, they didn't analyze the reason or propose strategies to solve the problem of oversupply. Kash *et al.* [15] demonstrate that there are significant disparities in the cost of new and old files in a private community named DIME, and users compensate for the high cost of older files by downloading more copies of newer files or by preferentially consuming older

files during free-leech periods. Particularly, they have shown that after a period of free-leech, there are more download activities in the community. This is consistent with our result that during free-leech, there is more demand and the oversupply is alleviated. Besides analyzing positive and negative effects of SRE, we also extensively explore the user behaviors and argue the reasons for these behaviors. Further, we analyze the performance of well-adopted community strategies, their effects against strategic user behavior, and the remedies we proposed.

9. Conclusion

While previous work only focuses on showing the effectiveness of SRE in incentivizing users to contribute, in this paper we provide a better understanding of private communities by exploring the user behaviors and demonstrating both the positive and the negative effects of these behaviors. We show that swarms in private communities are greatly oversupplied. Users achieve very high download speeds, but they need to seed for excessively long times to maintain required sharing ratios. Moreover, their seedings are often not very productive (in terms of low upload speed) and that their long seeding times do not necessarily lead to large upload amounts. Meanwhile, as users evolve in the community some users become more committed, in terms of higher ratios of the seeding and the leeching time, and some users game the system by keeping risky low sharing ratios while they leech more often than they seed. For users who intend to increase their sharing ratios, we show that seeding for longer durations is not as effective as increasing the upload speed. If it is not realistic for the users to upgrade their internet access, we suggest them to join swarms early or to join undersupplied swarms.

References

- [1] X. Chen, X. Chu, Measurements, analysis and modeling of private trackers, in: Proceedings of the 10th IEEE International Conference on Peer-to-Peer Computing (P2P'10), 2010.
- [2] Z. Liu, P. Dhungel, D. Wu, C. Zhang, K. Ross, Understanding and improving incentives in private P2P communities, in: Proceedings of the 30th International Conference on Distributed Computing Systems (ICDCS'10), 2010.
- [3] M. Meulpolder, L. D'Acunto, M. Capotă, M. Wojciechowski, J. Pouwelse, D. Epema, H. Sips, Public and private BitTorrent communities: A measurement study, in: the 9th International Workshop on Peer-to-Peer Systems (IPTPS'10), 2010.
- [4] C. Zhang, P. Dhungel, Z. L. Di Wu, K. Ross, BitTorrent darknets, in: Proceedings of the 30th IEEE International Conference on Computer Communications (INFOCOM'10), 2010.

- [5] A. Jia, X. Chen, X. Chu, J. Pouwelse, D. Epema., How to survive and thrive in a private BitTorrent community, in: the 14th International Conference on Distributed Computing and Networking (ICDCN'13), 2013.
- [6] R. Rahman, D. Hales, T. Vinkó, J. Pouwelse, H. Sips, No more crash or crunch: Sustainable credit dynamics in a P2P community, in: Proceeding of International Conference on High Performance Computing & Simulation (HPCS'10), 2010.
- [7] A. Jia, R. Rahman, T. Vinkó, J. Pouwelse, D. Epema, Fast download but eternal seeding: The reward and punishment of sharing ratio enforcement, in: Proceedings of the 11th International Conference on Peer-to-Peer Computing (P2P'11).
- [8] A. Jia, R. Rahman, T. Vinkó, J. Pouwelse, D. Epema, Systemic risk and user-level performance in private P2P communities, IEEE Transactions on Parallel and Distributed Systems Available online.
- [9] CHDBits, <http://chdbits.org/>.
- [10] HDStar, <http://www.hdstar.org/>.
- [11] ChinaHDTV, <http://www.chinahdvtv.org>.
- [12] A. L. Jia, X. Chen, X. Chu, J. Pouwelse, From user experience to strategies: how to survive in a private community, Tech. Rep. PDS-2011-004, Delft University of Technology (2011).
- [13] N. Andrade, E. Santos-Neto, F. Brasileiro, M. Ripeanu, Resource demand and supply in BitTorrent content-sharing communities, Computer Networks 53(4):515-527.
- [14] C. Spearman, The proof and measurement of association between two things, American Journal of Psychology 15:72-101.
- [15] I. Kash, J. Lai, H. Zhang, A. Zohar, Economics of BitTorrent communities, in: the 6th Workshop on the Economics of Networks, Systems, and Computation (NetEcon'11), 2011.
- [16] L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, X. Zhang, Measurements, analysis, and modeling of BitTorrent-like systems, in: Proceedings of the 5th Internet Measurement Conference (IMC'05), 2005.
- [17] J. Pouwelse, P. Garbacki, D. Epema, H. Sips, The BitTorrent P2P file-sharing system: Measurements and analysis, in: the 4th International Workshop on Peer-to-Peer Systems (IPTPS'05), 2005.
- [18] M. Meulpolder, L. Meester, D. Epema, The problem of upload competition in Peer-to-Peer systems with incentive mechanisms, Concurrency and Computation: Practice and Experience. Available online.

- [19] BitSoup, <http://bitsoup.org>.
- [20] PolishTracker, <http://polishtracker.net/>.
- [21] R. Cuevas, M. Kryczka, A. Cuevas, S. Kaune, C. Guerrero, R. Rejaie, Is content publishing in BitTorrent altruistic or profit-driven?, in: the 6th International Conference on emerging Networking EXperiments and Technologies (CoNEXT'10), 2010.
- [22] X. Chen, X. Chu, Z. Li, Improving sustainability of private P2P communities, in: Proceeding of the International Conference on Computer Communications and Networks (ICCCN'11), 2011.