Measurements, Analysis and Modeling of Private Trackers

Xiaowei Chen
Computer Science Department
Hong Kong Baptist University
Hong Kong
xwchen@comp.hkbu.edu.hk

Yixin Jiang
Computer Science Department
Tsinghua University
Beijing, China
yixin.tsinghua@gmail.com

Xiaowen Chu
Computer Science Department
Hong Kong Baptist University
Hong Kong
chxw@comp.hkbu.edu.hk

***Abstract*—BitTorrent plays a very important role in the current Internet content distribution. The enormous impact of public and private trackers should not be overlooked. Public trackers are suffering from free-riding problem, but private trackers are becoming more and more popular and they run very well in terms of an effective Share Ratio Enforcement (SRE) which is an auxiliary incentive mechanism. In this paper, we have crawled and traced 15 trackers with 3.5 million torrents for over 6 months. We first provide taxonomy of private trackers, and then present in breadth and depth measurement from the user viscosity, torrents evolution, user behaviors, content distribution and other metrics. Some features are apparently different from public trackers. Furthermore, we analyze SRE mechanism and point/credit system, and use game theory to study the effectiveness of SRE. There exists "uploading starvation" phenomenon in private trackers. We model SRE mechanism and preliminary propose an improved SRE mechanism to further incent users and enhance the performance of private trackers.**

Keywords-private tracker; incentive mechanism, BitTorrent, content distribution, peer-to-peer networks

I. INTRODUCTION

BitTorrent is currently the dominating P2P file sharing protocol. As one of the core components in BitTorrent protocol, trackers play an important role during the distribution process which periodically provides updated peer lists to connected clients. BitTorrent trackers can be divided into two categories, public trackers and private trackers [1]. Public trackers can be used by anyone by adding some tracker addresses to an existing torrent. There are many public trackers such as the Pirate Bay, Mininova, ISOHunt, etc. Although BitTorrent has implemented the Tit-for-Tat (TFT) algorithm as an incentive mechanism, public trackers are still suffering from free-riding problem: first of all, a peer may stop uploading immediately after it finishes the download task; secondly, a peer usually sets a limit on the total upload bandwidth.

In recent years, private trackers (PTs, also known as private tracker sites, darknets, etc.) become more and more popular and generate a huge amount of Internet traffic [2]. E.g., Rutracker.org (a.k.a. Torrents.ru) alone has more than 1.07PB of content (nearly 824,000 torrents). It can generate more than 60GB/s of Internet traffic. Users in PTs can usually achieve much faster download speed than they are in public trackers. Based on strict member controlling policy, PTs adopt Share Ratio Enforcement (SRE) as an auxiliary incentive mechanism to overcome the free-riding issue. SRE forces registered users to maintain a share ratio (i.e., upload-to-download ratio). In general, a registered user will be banned from PT community if his share ratio is lower than a threshold (e.g., 0.8). In addition, a user with higher share ratio will be awarded. There are two ways to achieve a high share ratio: one is to provide a high upload bandwidth; another one is to prolong the seeding time.

The main differences between public trackers and private trackers are summarized as follows:

- Users & Torrents Scale: Public trackers are open to everyone, but PTs are only available for registered users. It is remarkable that the number of torrents in PTs collectively will be much more than that in public trackers in existence [3].
- Incentive Mechanism: Public trackers rely on the BitTorrent TFT algorithm. PTs implement an additional SRE mechanism to incent the users to contribute as much as possible.
- Traffic Counting: Public trackers do not count each user's traffic during content distribution. PTs accurately record each user's total amount of uploaded data (T_u) and downloaded data (T_d). A user's ***Share Ratio*** is then calculated as T_u/T_d. If a user's Share Ratio is lower than a predefined threshold, he will be banned from this PT community.
- Download Performance: The download performance of PTs is usually much better than public trackers due to the SRE mechanism. This is because a user in PT is incented to provide a high upload bandwidth and a long seeding time.

Given the importance and popularity of PTs, it is essential to understand their characteristics so as to help us design better mechanism and build better sustainable environment to BitTorrent content distribution. We conduct a measurement study and provide a thorough analysis on the collected datasets. Our main contributions are summarized as follows:

- We have crawled 13 private trackers and 2 public trackers from Sep. 28, 2009 to Apr. 10, 2010. We obtained 31 datasets that cover 3.5 million torrents.

This work was supported by grant FRG2/08-09/098 and FRG2/09-10/081 from Hong Kong Baptist University.

- By analyzing the collected datasets, we investigate private trackers from the user viscosity, torrents evolution, user behaviors and content distribution aspects in detail. Some features are apparently different from public trackers. We show that SRE is an effective mechanism to incent users to seed as much as possible. But SRE will lead to "uploading starving" issue.
- Furthermore, we analyze SRE and auxiliary credit system, and use game theory to study effectiveness of SRE mechanism. We model the mechanism and preliminary propose an improved SRE mechanism to enhance the performance of private trackers.

The rest of our paper is organized as follows. Section II provides PTs' overview. Section III describes our measurement methodology and the collected datasets. Section IV presents detailed analysis. The modeling of SRE based on game theory is presented in section V. Related work is presented in section VI, following by our conclusions in Section VII.

II. Overview of Private Trackers

A. Taxonomy Overview

We classify the roles and operation processes of PTs. The roles in PT are divided into the two parts shown in Table I. In general, PT's sysops establish and publish PT, and then operate it by attracting users with high ranking [4]. Users sign up in PTs by using some methods [5]. After that, registered users/members select accepted clients to seed or leech contents above minimum share ratio, and then contribute if possible (e.g. uploading, donation, etc.). For each role and operation process we give related terms and definitions in part B.

TABLE I. Breakdown of Private Trackers

Private Trackers		
Role	***Operation Process***	***Related Terms***
Tracker User	Registration	Invitation Code
	Sharing (Seeding & Leeching)	Share Ratio, Passkey, User Class, Snatched, Freeleech, HnR,
	Contribution	Credit/Point System
Tracker Owner	Building & Publishing	Codebase
	Ranking Improving	Scene Release

"Related Terms" column does not cover all components to operation processes.

B. Terminologies and Definitions

The terms used in the PTs are not standardized. For the sake of clarity, we try to define the terms shown in Table I, which are widely used in PTs. Note that the terms list here are non-exhaustive.

Invitation Code: It is a typical registration way to become a member of closed PT communities. Contributing members meet specific requirements are eligible to invite friends to join PT by sending an invitation code.

Share Ratio: The PT calculates the share ratio for each user, which is the total amount of data the user has uploaded, divided by the total amount it has downloaded.

Passkey: It is a unique identity that PTs assign each registered user. Passkey is usually a hexadecimal string. It is appended to the announcing URL in the .torrent file which is dynamically generated for each member. This is to prevent private torrents from being uploaded to public websites. Well behaved users should not leak the passkey and announce URL to other members or public trackers; otherwise, they will be banned from the PTs.

User Class: Most PTs deploy a ranking system to categorize their users based on each user's contributions to the community. For example, the users can be categorized from low to high rank as newcomer, user, power user, elite user, VIP, etc. User's contributions normally are the downloaded and uploaded volume, share ratio, time of being a register user, etc. Different user classes have different privileges.

Snatched: It indicates that how many times a torrent file is completely downloaded.

Freeleech: When a torrent is flagged with freeleech, it means leeching that torrent will neither affect member's download amount nor decrease his share ratio. Meanwhile seeding that torrent will continue to increase upload amount.

HnR: It refers that a peer stops uploading as soon as it completes the downloading. HnR are highly prohibited. Different trackers have different rules about HnR.

Credit/Point System: Many PTs incorporate a "credit/point system" with SRE mechanism. Points can be earned maintaining a good share ratio; uploading torrents, etc. Points can be spent to improve member's account authority.

Codebase: Typically PTs do not develop sites, but instead use or modify some open source codebases.

Scene Release: Basically it means that content which has been first released by someone or some groups.

C. Operation Principle of Private Tracker

PTs implement a strict set of rules to control member eligibility and content quality. Users sign up to be a member of PT through an invitation system. Under a passkey system, each member is given a unique announcing URL to perform content distribution. In a predefined User Class system which adopts SRE mechanism, members can be automatically or manually promoted or demoted. The typical operation principle of PT is illustrated in Fig. 1.

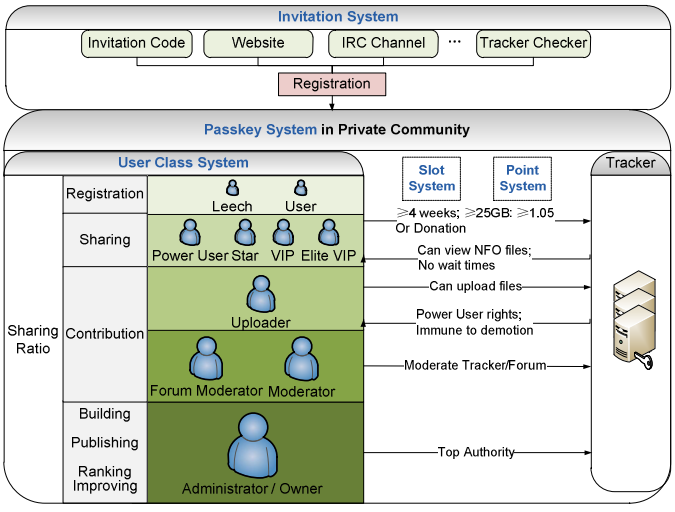

Figure 1. The Typical Operation Principle of Private Tracker

III. MEASUREMENT SETUP

A. Selection of Private Trackers & Collection of Data

There are a huge amount of PTs on the Internet [7]. Most of them are closed systems, and it is very difficult for an outsider to receive an invitation code. Nevertheless, based on the list and categories of [4] and [7], we successfully joined 13 representative PTs to perform our measurement study. For comparison purpose, we also crawled 2 well-known public trackers. We summarize all these trackers in Table II.

TABLE II. TRACKER LIST

Category		Tracker Information		
		Name	***# of Active Torrents***	***# of Registered Users***
Public	General	thePirateBay	2,701,080	4,347,019
		TorrentPortal	2,456,412	1,208,876
Private	General	Demonoid	271,095	N/A
		ILoveTorrents	8,886	N/A
	Scene Release	TorrentLeech	26,119	N/A
		RevolutionTT	32,064	N/A
		Bitsoup	12,955	N/A
	HD (High Definition)	CHDBits	16,317	25,378
		HDStar	11,792	8,172
		HD-Torrents	14,837	29,113
	Foreign	RuTracker.org	823,918	4,859,168
	Music	DimeaDozen	37,909	110,327
	TV	TheBox.bz	55,030	N/A
	DVD	AsianDVDClub	27,996	54,077
	Adult	PureTNA	72,778	1,335,367

Data updated until 20:00, April 10, 2010. N/A indicates the tracker does not provide it.

We use passive method to collect PT traces by using a crawler developed by ourselves. Data collection is recorded periodically by crawling web pages provided by each PT. These datasets are divided into three aspects, as shown in Table III. "Tracker statistics" covers the information of registered users, torrents, seeders, leechers, and seeder-to-leecher ratio, etc. Table IV provides "tracker statistics" in detail. "Top 10" includes information about top members/torrents based on traffic, speed, share ratio, etc. "Torrent list" contains detailed torrent information, such as torrent type, name, size, added time, snatched times, seeders, leechers, etc. Notice that trace duration includes tracker maintenance periods during which our crawling cannot be done. In addition, we cannot start crawling at the same time to all the trackers due to the difficulty of getting invitation code or open registration time.

B. Challenging Issues during Crawling

There are two challenging issues we need to address: incomplete torrent list of public trackers, information loss caused by crawling time intervals.

ThePirateBay only allows displaying recent 50 pages for each torrent category and 300 pages for all torrents, so we cannot depict the whole picture of thePirateBay, but we have crawled as many torrents as we can. On the other side, we have successfully collected the complete torrent lists from all PTs.

The frequency of data collection should be carefully tailored to achieve a smooth crawling. Many trackers enforce a minimum time interval between two page visits. To solve this issue, we estimate the relation between information loss ratio with time interval. We configure our crawlers to execute every 4 hours for tracker statistics and torrent list, every day for Top 10 because its change is not frequent. Complete torrent list can reflect the overall population of torrents in the tracker. Furthermore, we have crawled two popular torrents every 10 minutes to study torrent life span.

TABLE III. CRAWLING DATASETS

Tracker Name	Tracker Statistics	Top 10	Torrent List	Trace Duration (days, mm/dd/yy)
thePirateBay	√	N/A	(Part)	103, 12/29/09 – 04/10/10
TorrentPortal	√	N/A	√	67, 02/02/10 – 04/10/10
Demonoid	N/A	N/A	√	84, 01/17/10 – 04/10/10
ILoveTorrents	N/A	N/A	√	193, 09/28/09 – 04/10/10
TorrentLeech	N/A	Only Users	√	86, 01/15/10 – 04/10/10
RevolutionTT	N/A	N/A	√	96, 12/26/09 – 04/10/10
Bitsoup	N/A	N/A	√	177, 10/16/09 – 04/10/10
CHDBits	√	Top 250	√	193, 09/28/09 – 04/10/10
HDStar	√	√	√	186, 10/05/09 – 04/10/10
HD-Torrents	√	N/A	√	78, 01/23/10 – 04/10/10
RuTracker.org	√	N/A	N/A	190, 10/01/09 – 04/10/10
DimeaDozen	√	Speed N/A	√	68, 02/02/10 – 04/10/10
TheBox.bz	N/A	N/A	√	68, 02/02/10 – 04/10/10
AsianDVDClub	√	√	√	68, 02/02/10 – 04/10/10
PureTNA	√	Only Torrents	√	68, 02/02/10 – 04/10/10

N/A indicates the tracker does not provide it or we cannot acquire it.
Only DimeaDozen provides the statistics of each torrent's transfer speed in Torrent List.

TABLE IV. ELEMENTS OF TRACKER STATISTICS

Tracker Name	Tracker Statistics				
	# of Peers, Torrents, Users	Traffic	User Class	Seeder /Leecher Ratio	Active browsing & tracker users
thePirateBay	√	N/A	N/A	N/A	N/A
TorrentPortal	√	N/A	N/A	N/A	N/A
CHDBits	√	√	√	√	√
HDStar	√	√	√	√	√
HD-Torrents	√	√	N/A	√	N/A
RuTracker.org	√	√	N/A	N/A	N/A
DimeaDozen	√	√	√	√	N/A
AsianDVDClub	√	N/A	N/A	√	N/A
PureTNA	√	N/A	N/A	√	N/A

Traffic generally includes total uploaded and total downloaded data.
RuTracker.org & DimeaDozen **ONLY** provide data transfer rate in Traffic.

IV. MEASUREMENT AND ANALYSIS

We will present detailed measurement and analysis based on the statistics in crawled datasets. We assume that PTs have global, trustful and accurate view of torrent's information (e.g., upload and download volume). We studied the user viscosity, torrent evolution, user behaviors, distribution of private trackers, and the comparison with public trackers.

A. User Viscosity of Private Trackers

According to the traffic ranking from Alexa [9], we select 6 popular public trackers (thePirateBay.org, Torrentz.com, ISOHunt.com, BTJunkie.org, TorrentReactor.net, Mininova.org) and 6 famous PTs (Torrents.ru, Demonoid.com, TorrentLeech.org, PureTNA.com, Bitsoup.org, TheBox.bz) from our crawling list.

Page views measure the number of pages viewed by site visitors. Multiple page views of the same page made by the same user on the same day are counted only once. Page Views

per User is the average number of unique pages viewed per user per day. Fig. 2 shows the current page views per user of each tracker from Alexa. It is obvious that PTs have higher page views per user than public trackers. This implies that users in PTs have more interest than public trackers. User viscosity of PTs is better than public trackers.

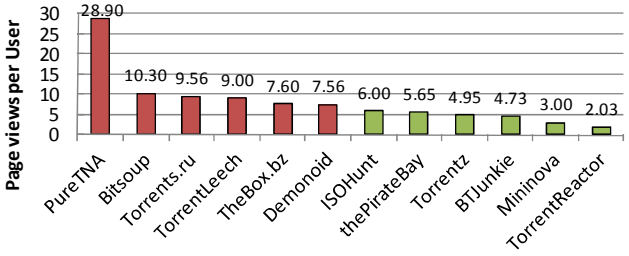

Figure 2. Page views per User in Popular Trackers (until Feb. 6, 2010)

Next, we take a close look at this viscosity of users by analyzing the number of users in each class of CHDBits. Fig. 3 shows our measured data of CHDBits in a 3-month period[1]. The total number of users keeps relatively stable[2]. In general, stability of total number of users is applicable to other PTs. The stability is mainly depended on two factors: the capacity of the PT and the maintenance of the tracker's administrator.

Different PTs adopt different user class policy, but their internal structures are very similar. In Fig. 3, most of the users belong to the User class (i.e. downloaded over 50GB and share ratio is lower than 0.95). We observe that at the number of users in User class in decreasing with time, while the numbers of users in Power User and Elite User classes are increasing. It shows the effectiveness of the ranking system in PTs as users are incented to promote themselves to a higher user class.

Though the number of members keeps stable, but the most members are moving to higher level, shown In Fig. 3. The reason of this phenomenon is that sysops provide freeleech occasionally in order to motivate members to increase their share ratios. On one hand, this method helps members keep staying in the PT. On the other hand, it stimulates members to be active in PTs. Keeping member as active as possible is essential in PT, which is discussed in subsection C.

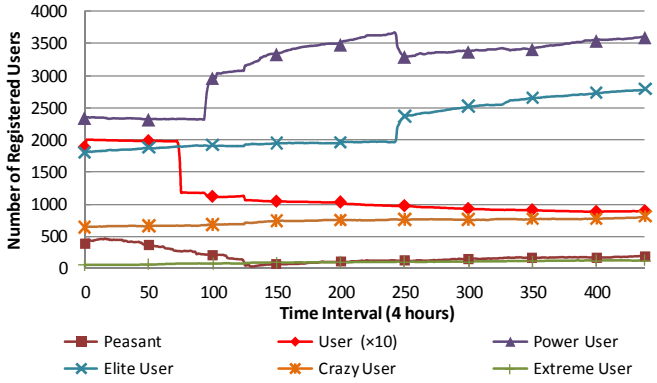

Figure 3. Number of Registered Users of each User Class in CHDBits

[1] Notice that the unit of curves "User" is 10.

[2] The reason of the big drop of User class around time 70 is because the administrator deleted many inactive members

B. *Torrents of Private Trackers*

In this subsection we investigate the behavior of a single torrent and compare with a torrent in public tracker TorrentPortal. Fig. 4 presents the evolution of torrents about two movies (same content with different resolutions) based on the number of seeders, leechers and snatched times in CHDBits. The main finding in Fig. 4 is that the number of seeders is significantly larger than the number of lechers most of the time. Other PTs show the similar features [10].

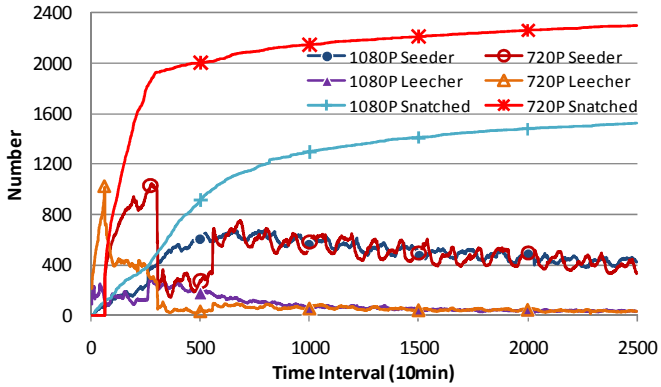

Figure 4. Seeder, Leecher, Snatched times of Two Torrents in CHDBits

Fig. 5 shows the number of peers in an active torrent (i.e., a torrent with at least one seeder) between PTs (different categories) and one public tracker. Torrents are ordered from the largest to smallest in terms of the number of seeding peers in a torrent. X-axis's label denotes torrent sequence number. We can see from Fig. 5 that active torrent ratio in public tracker TorrentPortal drops down the quickest than all of PTs. It indicates that torrents are more active in PTs than in the public tracker. That guarantees the high download speed in PTs. Besides, we can see there exist "giant" torrents not only in public trackers, but also in PTs which contain more than 1,000 seeders. Of course, due to the limitation of server capacity, the "giant" scale in PT cannot reach the level of public tracker. E.g. in TorrentPortal, there are ten torrents with 8,388,607 seeders, and some of them even have 8,388,607 leechers. These torrents are usually popular TV series and movies.

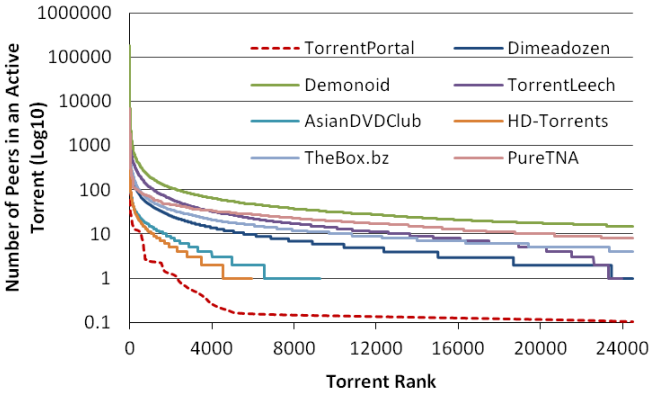

Figure 5. Active Torrent Distribution by the Number of Seeding Peers

We select PT (DimeaDozen) to take a close look about age of torrents distribution and transfer speed of torrents

respectively (Fig. 6 and Fig. 7). Note that only DimeaDozen provides the statistics of each torrent's transfer speed. In Fig. 6, about 40% torrents are uploaded in 3 months and 82% of torrents are uploaded within one year. Surprisingly, some torrents with nearly 3 years are still active. The other PTs represent the similar results. In Fig. 7, though there are a lot of long live torrents with seeders in PT, more than 98% torrents with small number of leechers have no more than 40-50KBps in this music PT, only those popular torrents attract many leechers and have very high speed. It implies PT has "***uploading starving***" condition that the number of seeders is significantly larger than the number of leechers. Furthermore, because the supplies from seeders exceed the demands of leechers, so the leechers' download speed can be very high in PTs. Uploading starving will be discussed in section V.

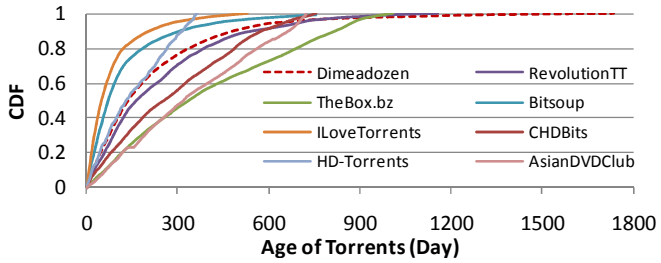

Figure 6. Age Distribution of Torrents

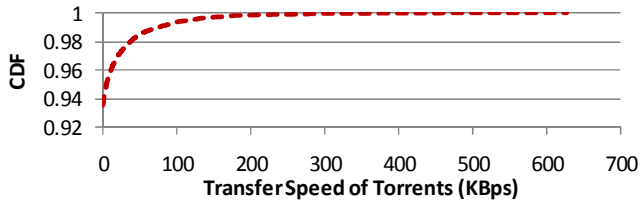

Figure 7. Transfer Speed Distribution of Torrents in DimeaDozen

C. User Behaviors of Private Trackers

This subsection presents our findings about user behaviors in PTs. From section III, we know that only CHDBits and HDStar provide detailed statistics of active browsing users and active tracker users.

1) Activity of Users

Active user is a set which includes two parts: active tracker user means he is joining the BitTorrent content distribution (uploading/downloading), while active browsing user means he is only browsing the PT site. Fig. 8 shows the active user ratio (active users to total users) statistics per day and per week in CHDBits and HDStar. In Fig. 8, though these two trackers have different number of users, their active user ratios are surprisingly very close to each other: the per day ratio keeps around 50% and per week ratio keeps around 80%. From Fig. 8 we can see that, compared with active user ratio per week, there still exists nearly 30% members who are inactive per day.

Then we select one week period (Oct. 13, 2009 to Oct. 20, 2009) in CHDBits and HDStar to compare the change of active browsing user and active users connecting to the tracker (for short, active tracker users). In Fig. 9, the active browsing users in two trackers present the same trend with the change of time. The wave peak happens at 22:00PM every day and the sub-peak is at 10:00AM every day, which means that there are most of browsing users around these two time points. The wave hollow is located at 6:00AM every day.

When adding the factor of active tracker users, the result of ratio (active tracker users to active browsing users) in Fig. 10 shows that the wave peak becomes 6:00AM, while the wave hollows are at 10:00AM and 22:00PM. That means most users in PTs keep staying in the BitTorrent swarm. For example, at 6:00AM, most users are not browsing, but their clients keep connecting to the tracker, so the ratio of active tracker users to active browsing users presents peak status. The reason of the CHDBits irregular wave changes (around 22:00PM, Oct. 17, 2009 to 10:00AM, Oct. 18, 2009) in Fig. 9 and Fig. 10 is that the tracker suffered non-periodical DDoS attack, which also affects the tracker in the following 2 days, so members' clients have difficulty in connecting to the tracker.

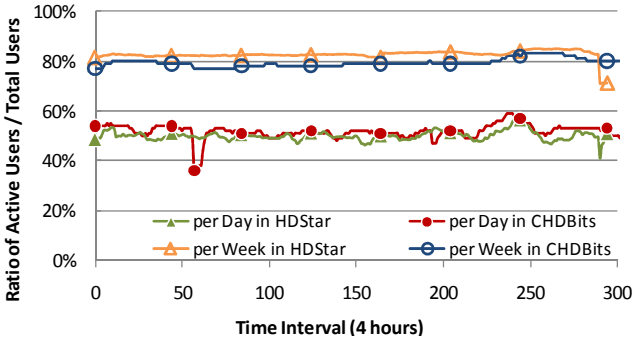

Figure 8. Ratio of Active Users/Total Users per Day & per Week

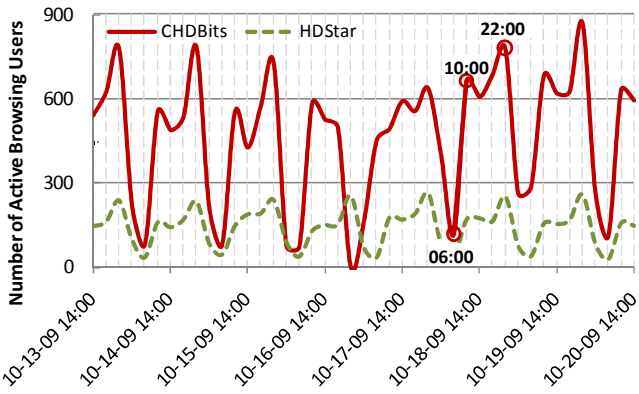

Figure 9. Number of Active Browsing Users

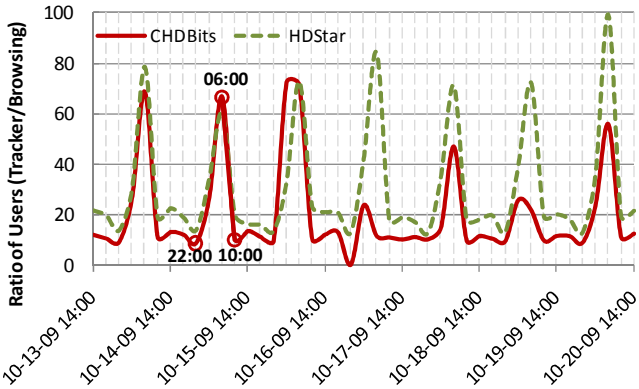

Figure 10. Ratio of Tracker Active Users/Active Browsing Users

2) Activity of Peers

Fig. 11 shows the ratio of seeding peers to leeching peers in PTs. Generally, the number of seeding peers in PTs is 6 to 15 times of the number of leeching peers. Users in PTs are seeding more than leeching. The wave properties of trackers follow the interchange pattern of daytime and nighttime.

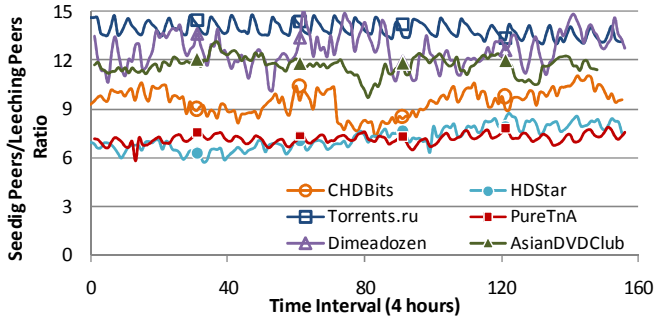

Figure 11. Ratio of Seeding Peers/Leeching Peers in Private Trackers

3) Traffic of Private Trackers

This subsection shows that statistics of the traffic of PTs. Traffic here indicates the sum of uploaded and downloaded traffic. In Fig. 12, on one hand, we can see that the total traffics per user of CHDBits and HDStar trackers keep increasing from 430TB, 496TB to around 955TB, 851TB, respectively. Fig. 13 shows that total traffic in Torrents.ru is surprising 15-35GBps with nearly 700,000 living torrents. The tremendous traffic may bring heavy burden to our ISPs. On the other hand, the ratios (total uploaded to total downloaded) of two trackers maintain the same value (around 5). This further proved that users in PTs are doing more seeding than leeching. It reflects a kind of imbalance: supply is more than demand, which also proves that the high download speed in PT is partly because of the enormous seeders in the swarm.

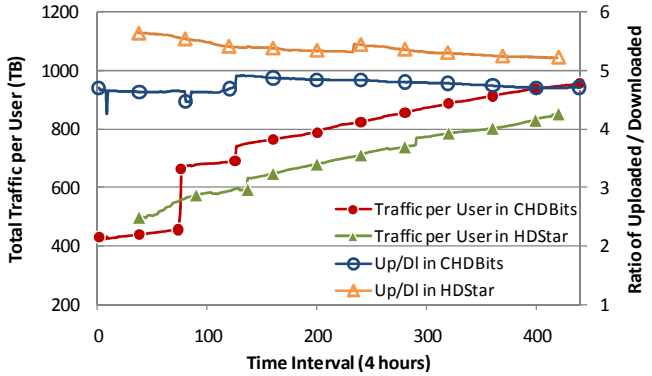

Figure 12. Total Traffic/User & Ratio of Total Uploaded/ Downloaded

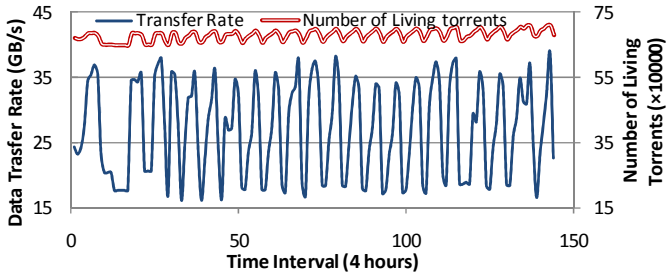

Figure 13. Data Transfer Rate in Torrents.ru

4) Top 250 & Top 10 in Private Trackers

According Table III, different PTs provide different top 10 lists. We select PTs who have common features. Only CHDBits provides lists of Top 250, such as best downloaders, best uploaders, best share ratios, etc.

Fig. 14 shows the change of top 250 members with best share ratios. Though CHDBits have over 15,000 members, there are no more than 250 members who have share ratios of over 10.

Actually, most of members only have the share ratio of no larger than 1. E.g., in DimeaDozen, there are 39% members whose share ratios are lower than 0.25, 75% members whose share ratios are lower than 1.

Fig. 15 presents the fastest uploading and downloading speed (including inactive time) in top 250 members. Normally, average uploading speed is lower than downloading speed in xDSL broadband configuration. The opposite situation in Fig. 15 represents that a lot of members keep seeding for a long time but they choose to leech from time to time. Therefore, if calculated average speed with their inactive time, the average speed of uploading is greater than downloading.

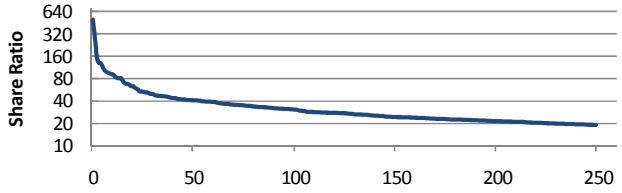

Figure 14. Top 250 Members with Best Share Ratio in CHDBits

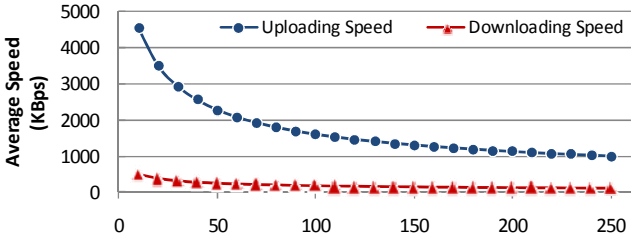

Figure 15. Top 250 Members with Fastest Transfer Speed in CHDBits (including inactive time)

Fig. 16 shows the top 10 members with best share ratio, which result is similar with Fig. 14. Fig. 17 shows the top 10 torrents with best share ratio. Fig. 18 shows the top 10 members with fastest transfer speed (including inactive time), which result is similar with Fig. 15.

From Fig. 16 and Fig. 17, we may deduce that small number of members accounts for seeding large number of torrents.

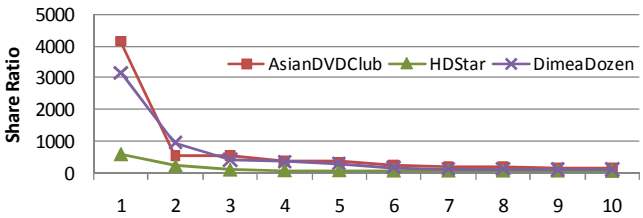

Figure 16. Top 10 Members with Best Share Ratio

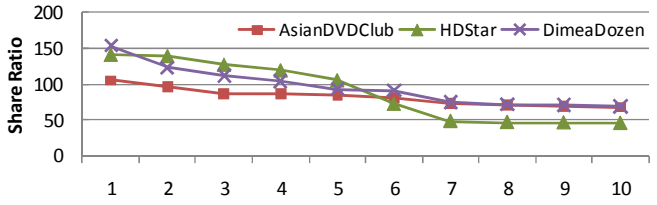

Figure 17. Top 10 Torrents with Best Share Ratio

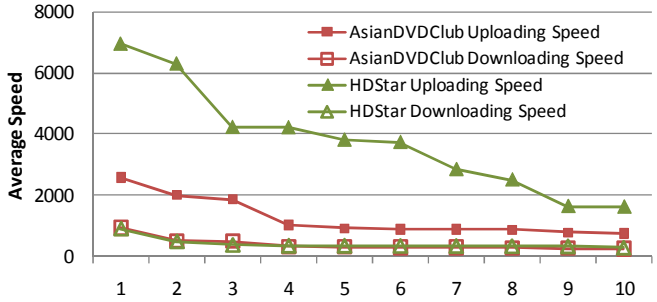

Figure 18. Top 10 Members with Fastest Transfer Speed (including inactive time)

5) Active Torrents Rate of Private Trackers

Active torrents rate is the ratio of the number of torrents which have at least one seed to the total number of torrents in the tracker. Along with the increasing torrents scale, maintaining the active torrents rate is one of KPIs to evaluate the active level and ranking of trackers. Fig. 19 compares the active torrents rate of 14 trackers. Apparently, most PTs are far better than public trackers in terms of active torrents rate.

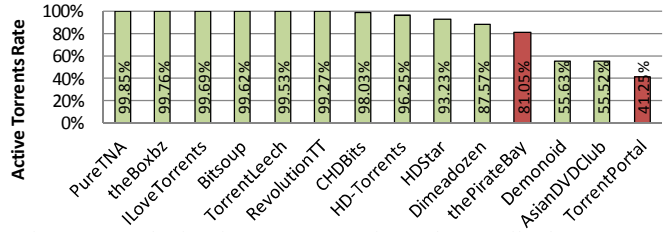

Figure 19. Rank of Active Torrents Rate in Trackers (until Feb. 6, 2010)

D. Content Distribution of Private Trackers

This subsection shows the content size distribution and category distribution in PTs. In order to be representative, we show results of one music tracker (DimeaDozen), one HD tracker (CHDBits) and one general tracker (RevolutionTT).

1) Content Size Distribution

Fig. 20 shows the classification of content size distribution based on the number of contents in two PTs, respectively. We can see the most number of contents are small contents with 0-1GB, followed by the contents which are large than 100GB.

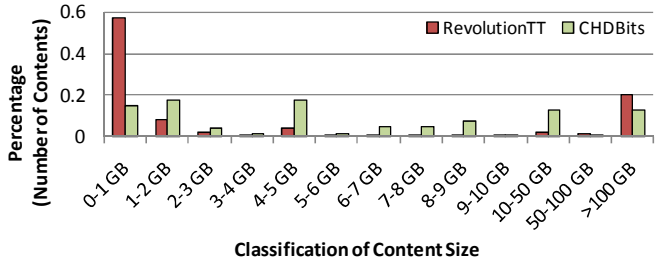

Figure 20. Distribution of the Number of Contents

2) Category Distribution

Fig. 21 and Fig. 22 show the category of content distribution based on the number of contents in two PTs, compared with peer distribution respectively. We can see that popularity of content category basically matches the peer distribution accordingly. The DimeaDozen provides detailed categories of music, that is, 61 categories managed strictly by private sysops, which shows a long tail distribution. We list the top 5 category percentage in Table V. Fig. 22 shows that TV series, movie and adult are the leading types which occupy 81% share in PT. There is a great diversity in the content distribution based on category.

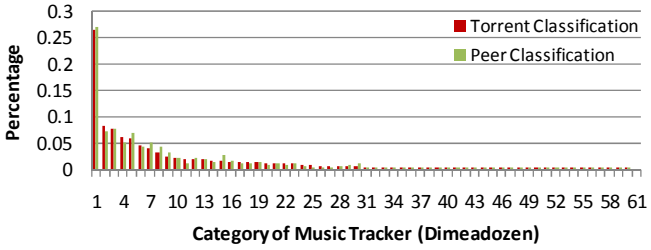

Figure 21. Classification of Peers and Torrents in Dimeadozen

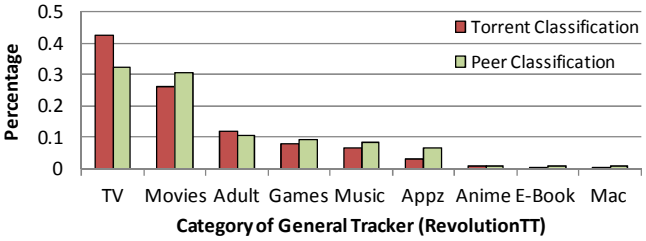

Figure 22. Classification of Peers and Torrents in RevolutionTT

TABLE V. TOP 5 CATEGORIES IN DIMEADOZEN

Category	Rock	Alternate	Jazz	Singer or Songwriter	Progressive Rock
Percentage	27.03%	7.17%	7.63%	5.15%	6.92%

Except for the above measurement, there are other metrics which can measure the performance of PTs, such as pre-time, firewall rate, freeleech, download slots, etc [10].

V. ANALYSIS AND IMPROVEMENT OF SRE MECHANISM

A. Analysis of SRE Mechanism

Before our analysis, we define several symbols, shown in Table VI.

TABLE VI. SYMBOLS DEFINITIONS IN SRE MECHANISM

Symbol	**Definition**	**Range**
n	uploaded/downloaded times. One time can be viewed as per torrent, or per day, etc.	N
U_n	the total amount of uploaded.	$R^+ \cup 0$
u	the amount of uploaded at each time i, i.e. $u_1, u_2, .., u_n$.	
D_n	the total amount of downloaded.	
d	the amount of downloaded at each time i, i.e. $d_1, d_2, .., d_n$.	
R	share ratio.	
r	share ratio during each time, i.e. $r_1, r_2, .., r_n$.	
Δ_n	the total amount of seeded by torrent provider.	
δ	the amount of seeded at each time i, i.e. $\delta_1, \delta_2, .., \delta_n$.	

In a closed private site system, we assume that no member contributes new torrent in the system, a member's share ratio is calculated as (1):

$$R_n = \frac{U_n}{D_n} = \frac{u_1+u_2+\cdots+u_n}{d_1+d_2+\cdots+d_n} \quad (1)$$

A member's share ratio during *i*th time r_i is defined as (3):

$$r_i = \frac{u_i}{d_i},\ i \in N \quad (2)$$

If a peer downloaded a new torrent in the (*n*+1)th time, then the new share ratio is calculated as (3):

$$R_{n+1} = \frac{U_n+u_{n+1}}{D_n+d_{n+1}} \quad (3)$$

If we want to increase the share ratio after every new downloading, that is:

$$R_{n+1} > R_n \quad (4)$$

Combining the above (1), (3) and (4), we can get (5):

$$\frac{u_{n+1}}{d_{n+1}} > \frac{U_n}{D_n} \quad (5)$$

That is:

$$R_{n+1} > R_n => r_{n+1} > R_n \quad (6)$$

Similarly, we can get:

$$R_{n+1} = R_n => r_{n+1} = R_n \quad (7)$$
$$R_{n+1} < R_n => r_{n+1} < R_n \quad (8)$$

From (6), (7) and (8), we know that the share ratio *R* is affected by each time's share ratio *r*.

Now we assume members can upload new torrents to private site. Then the actual share ratio's equation should be modified as following:

$$R'_n = \frac{U_n+\Delta_n}{D_n} \quad (9)$$

where $\Delta_n = \delta_1 + \delta_2 + \cdots + \delta_n$

If a system follows the rule that share ratio has relatively balanced supply and demand, that is:

$$U_n = D_n \quad (10)$$

Then (9) is revised as:

$$R'_n = 1 + \frac{\Delta_n}{D_n} \quad (11)$$

Equation (11) indicates that even in a relatively balanced supply and demand closed private site, if there are new torrents continue to be injected into the system by a member, Δ_n/D_n will continue to increase, so R'_n will increase as well. But most private sites have strict uploader management policies to select small part of "decent" members to be uploaders. So most of members' Δ_n/D_n are 0 and their R'_n should be 1 if they all could follow (10).

However, it is impossible that most of members' share ratios keep 1 which violates the law of supply and demand. Along with the increasing of some members' share ratios, the other members' share ratios must decrease at the same time.

Actually, there are a lot of members' ratios are lower than 0.95 ("User" rank in Fig. 3). This will generate an interesting phenomenon, "***uploading starving***", which means that a large number of members have to seed for a long time in order to increase their share ratios. Therefore, the number of seeding peers is 6 to 15 times of the number of leeching peers in the current private sites.

This status is relatively stable but unbalanced. In order to alleviate this imbalance of supply and demand, many private sites adopt freeleech promotion and credit system as additional incentive mechanisms.

B. Analysis of SRE Mechanism based on Game Theory

From a macroscopic view, the total download bandwidth is equal to the total upload bandwidth in a BitTorrent swarm. The reason of free-riding problem is that some users contribute a lot but receive unfair return. BitTorrent adopts TFT algorithm as a kind of incentive mechanism, but it has been shown that increased upload contribution only marginally improves download rates [11]. Peers lack the motivation to seed and have no reason to contribute once they have satisfied their immediate demands. In PTs, we have previously shown that SRE (Share Ratio Enforcement) is a very effective auxiliary incentive mechanism through our measurement study. Here we use a general game-theoretic framework to analyze SRE and answer the following question: Why does the TFT mechanism alone fail to achieve the same level of performance as PTs that use SRE?

In order to simplify the game, we assume that each peer has two strategy sets, upload (*UL*) and download (*DL*). N is the set of natural number.

$UL = \{(ul_1, ul_2, \ldots, ul_p) | p \in N\}$, denotes p different upload levels sorted from lowest to highest;

$DL = \{(dl_1, dl_2, \ldots, dl_q) | q \in N\}$, denotes q different download levels sorted from lowest to highest.

Each peer *k* selects a pair of (ul_i, dl_j) from *UL* and *DL*, and builds its own strategy space S_k, and then *n* peers in P2P system construct a strategy space set *S*

$$S_k = \{(ul_i^k, dl_j^k) | i, j, k \in N, 1 \le i \le p, 1 \le j \le q\}$$
$$S = \{(S_1, S_2, \ldots, S_n) |\ n \in \mathrm{N}\}$$

In general, peer wants to download by fulfilling its bandwidth configuration or maximum bandwidth. But it is difficult to download as pre-selected in *DL* set. In P2P content distribution network, the actual download performance is decided by the number of active peers joined in the swarm. Here we use the actual download performance of peer *i* to

define the utility u_i according to its own strategy s_i. The set of u is defined as:

$$u = \{(u_1, u_2, \dots, u_n) | \, n \in \mathrm{N}\}$$

Each peer has an optimal strategy s_i^* from its S_k. The optimal strategy set is defined as:

$$s^* = \{(s_1^*, s_2^*, \dots, s_n^*) | \, n \in \mathrm{N}\}$$

Then the optimal utility can be defined as:

$$u_i(s_1^*, s_2^*, \dots, s_i^*, \dots, s_n^*) \succ u_i(s_1^*, s_2^*, \dots, s_i', \dots, s_n^*),$$
where $\forall s_i' \in S_i, s_i' \neq s_i^*, i \in \mathrm{N}$

If peers are rational, they will choose to upload with minimum speed and download with maximum speed (ul_1, dl_q) as their optimal strategies. That is:

$$s^* = \{(ul_1^1, dl_q^1), (ul_1^2, dl_q^2), \dots, (ul_1^n, dl_q^n)) | \, n \in \mathrm{N}\} \quad (12)$$

If all peers upload with maximum speed, then peers' utility will become high according to TFT mechanism, that is:

$$s^{*\prime} = \{(ul_p^1, dl_q^1), (ul_p^2, dl_q^2), \dots, (ul_p^n, dl_q^n)) | \, n \in \mathrm{N}\} \quad (13)$$

From game theory, we know there are two equilibriums in this game: Nash equilibrium (12) and strict Nash equilibrium (13). Because peers are rational, so both sides will choose upload with minimum speed to reach the Nash equilibrium (12), and hence Tragedy of Common will occur. Though BitTorrent adopts TFT as a conflict resolution scheme, notice that TFT is based on repeated game and it can only take effect in infinite repeated game. However, there does not exist actual infinite repeated game but only finite repeated game in P2P content distribution. The result in finite repeated game is the same as in one time game [12]. So the final result of using TFT algorithm induces the Nash equilibrium (12).

But if we introduce SRE mechanism in BitTorrent content distribution system, peers cannot just be rational, and they are enforced to upload/seed to reach certain share ratio under a giving SRE policy. Even if a peer still chooses to upload with minimum speed, it must reach the required minimum share ratio by seeding for a long time. Therefore, the strict Nash equilibrium is broken, and the equilibrium will transit to the strict Nash equilibrium (13) which has Pareto efficiency. All peers download and upload with high speed. This will best utilize the potential of all peers and achieve the maximum efficiency of content distribution.

C. Modeling of SRE Mechanism

Here we choose one general tracker (RevolutionTT) and one HD tracker (CHDBits) to model their SRE mechanisms. Their SRE mechanisms are defined as Table VII. Other trackers' SRE mechanisms are similar [10].

In general, SRE mechanism is related with two variables: share ratio (r) and the volume of downloaded data (d). Though RevolutionTT and CHDBits have different SRE mechanisms, SRE in both trackers can be modeled as S-Curve: it first starts with a rapid growth stage and then enters into a stable and smooth stage. We model the S-Curve by the equation:

TABLE VII. SRE MECHANISM

Private Tracker	Downloaded Data	Share Ratio (no less than)
RevolutionTT	3.0GB	0.3
	6.0GB	0.50
	10.0GB	0.80
	>10.0GB	>0.8
CHDBits	10GB	0.3
	50GB	0.4
	100GB	0.5
	200GB	0.6
	400GB	0.7
	>400GB	0.95

In general, SRE mechanism is related with two variables: share ratio (r) and the volume of downloaded data (d). Though RevolutionTT and CHDBits have different SRE mechanisms, SRE in both trackers can be modeled as S-Curve: it first starts with a rapid growth stage and then enters into a stable and smooth stage. We model the S-Curve by the equation:

$$ln(r) = b_0 - \frac{b_1}{d}, \quad (14)$$
where $r \geq 0, d > 0, r, d \in R$

Through our regression analysis, b_0 is 0.179 and b_1 is -4.326 in RevolutionTT. In CHDBits, b_0 is 0.023 and b_1 is -11.81.

D. Credit System

When a torrent is flagged with freeleech, it means leeching that torrent will neither affect member's download amount nor decrease his share ratio. Meanwhile seeding that torrent will continue to increase upload amount. In general there are three freeleech promotion approaches: 30%, 50% or free, which means 30%, 50% or 0% of your download amount is counted respectively, but your upload amount is still counted 100%. From (9), we can see that this approach will definitely increase share ratio by increase the numerator (upload amount).

Credit system or point system is another approach deployed in private site to incorporate with SRE mechanism. Credit systems vary in different private sites, but their basic principles are nearly the same. A member's Credits can be spent to exchange his upload amount or invitation quota, etc. In general, the following method can gain more credits:

- Increasing the number of your seeding torrents
- Increasing uploading speed
- Increasing your seeding time
- Seeding more older torrents
- Seeding larger torrents

Actually the calculation method is complex in some PTs. In a word, credit system can increase share ratios of members, boost members' enthusiasm to seed more torrents for longer time, etc. Thus the private site stays more active.

E. Improvement of SRE Mechanism

Though the SRE mechanism achieves a big success in PTs, there still exist a lot of inactive members who are seldom active once they obtain high share ratio. They do harm to the sustainable development of PT communities. If there are too many inactive members with high share ratios, PTs cannot guarantee the number of seeders, and then the high download performance cannot be guaranteed either. Actually, some PTs changed its incentive SRE mechanism. E.g. Rutracker.org replaced SRE mechanism with time bonus points. Time bonus points are automatically gained when you seed downloaded torrents, regardless of the actual amount of data you upload back. It is good to low bandwidth users and those who struggled to maintain a healthy global ratio on this tracker.

We will consider the following factors to improve SRE:

- *R*: It is the share *R*atio that SRE used in PTs.
- *A*: It denotes the *A*ctivity of user in terms of seeding time $\Omega(\mathrm{t_s})$ and the number of seeding torrents $\psi(\mathrm{n_i})$.
- *P*: It denotes the *P*eriod of time to the last seeding $\mathrm{P(t_l)}$, the longer to the present, the larger the *P* is.

Our preliminary improved SRE mechanism (RAP) is shown as Eq. (15) to calculate the final ratio:

$$Ratio_{final} = r - P(\mathrm{t_l}) + \sum_{i=1}^{n} \Omega(\mathrm{t_s}) \times \psi(n_i), \qquad (15)$$

where $r, t_l, t_s, n_i \geq 0, r, t_l, t_s, \in R, n_i \in Z$

R, *A* and *P* can be recorded by PT server. Though peers could lie and manipulate such information, probability of these behaviors is very small. Once detected by PT server, user will be banned forever and even with his IP address. Note that getting invitation code of PTs is very difficult. Eq. (15) enforces that a user can not be inactive for a long time. The final ratio can be increased by normal methods, and also can be increased by the seeding time and number of seeding torrents.

VI. Related Work

To the best of our knowledge, there are limited studies based on measurements, analysis and modeling with BitTorrent PTs. In 2005, Nazareno et al firstly studied the Share Ratio Enforcement (SRE) in PTs [13]. They viewed the SRE as extrinsic gifting which is giving not motivated by a direct, immediate benefit [14]. Then Nazareno et al deeply studied another PT (bitsoup.org) with other two BitTorrent communities based on resource demand, supply and their relationship [8]. Until 2009, Tribler team proposed that if a peer uploads more than it downloads in a swarm means the peer should have more credit. PTs possibly exist "credit squeeze" which indicates that the lacking of credit leads to significantly reducing of system efficiency [15].

Though the previous investigations revealed some characteristics of PTs and obtained some primitive research results to SRE mechanism, the noticeable limitation is samples they studied are one or two PTs, which provides a limited view of current various PTs. It makes some ideas they proposed stayed in hypothesis stage but not conclusive. Recently C. Zhang et al. investigated 800+ PTs in terms of geography and content distribution. They compared PTs and public trackers comprehensively, contributed a lot about the difference between PTs and public trackers from macroscopic, medium-scopic and microscopic perspectives, especially in content overlapping, seed-to-leecher ratios and upload-to-download ratios. They found the aggregate size of PTs is very large. They studied one PT (HDChina) to investigate the influence of SRE mechanism. Above all, they depicted a broad and clear picture of PTs landscape [1]. Moreover, Z. Liu et al. studies in-depth SRE in PTs. They showed that PTs can provide more incentive for users to contribute and seed. They proved SRE's effectiveness with a game theoretic model and proposed an upload entropy scheme to prevent collusions in order to enhance robustness of PT sites [16].

VII. Conclusion

In this paper, we provided taxonomy of private trackers. We have crawled and compared system behaviors among 13 private trackers and 2 public trackers for over 6 months. We investigated private trackers from the user viscosity, torrents evolution, user behaviors, and content distribution in detail. In depth, we revealed the "uploading starving" phenomenon in PTs and studied the effectiveness of SRE mechanism based on game theory. We modeled the SRE and preliminary proposed an improved SRE mechanism. In summary, our study provided a comprehensively inside picture about private trackers.

References

[1] C. Zhang, P. Dhungel, Di Wu, Z. Liu,and Keith W. Ross, "BitTorrent Darknets," In Infocom 2010, San Diego, Mar. 2010.

[2] "The 100 Most Popular Private Trackers Of 2008," http://filesharefreak.com/2008/12/31/the-100-most-popular-private-trackers-of-2008/.

[3] "Private Trackers with the Most torrents," http://filesharefreak.com/2009/09/19/private-trackers-with-the-most-torrents/.

[4] "Torrentking.org," http://www.torrentking.org/?order=rating.

[5] "The Essential Guide to Getting Into Private Trackers," http://filesharefreak.com/2009/06/18/the-essential-guide-to-getting-into-private-trackers/.

[6] "CHDBits," http://chdbits.org/faq.php.

[7] "Trackers List," http://filesharefreak.com/trackers-list/.

[8] Nazareno Andrade, Elizeu Santos-Neto, Francisco Brasileiro, et al. "Resource demand and supply in BitTorrent content-sharing communities," Computer Networks Journal, Dec. 2008.

[9] "Alexa," http://www.alexa.com/.

[10] X. Chen, X.-W Chu, "Understanding Private Trackers in BitTorrent Systems" Technical Report, http://www.comp.hkbu.edu.hk/~xwchen/private_tracker/pt_tech_2010.pdf, Jul. 2010.

[11] M. Piatek, T. Isdal, A. Krishnamurthy, T. Anderson, "One hop reputations for peer to peer file sharing workloads," In NSDI 2008.

[12] Gibbons, R, "A Primer in Game Theory," Harvester, New York, 1992.

[13] Nazareno Andrade, Miranda Mowbray, Aliandro Lima, et al. "Influences on cooperation in BitTorrent communities." Third Workshop on the Economics of Peer-to-peer Systems, In SIGCOMM 2005, Aug. 2005.

[14] Matei Ripeanu, Miranda Mowbray, Nazareno Andrade, and Aliandro Lima, "Gifting Technologies: A BitTorrent Case Study," First Monday, vol. 11(11), Nov. 2006.

[15] D Hales, R Rahman, B Zhang, M Meulpolder, et al. "BitTorrent or BitCrunch: Evidence of a credit squeeze in BitTorrent?" In the Proc. of the 5th COPS Workshop, Jun. 2009.

[16] Z. Liu, P. Dhungel, D. Wu, C. Zhang, Keith W. Ross, "Understanding and Improving Incentives in Private P2P Communities," In ICDCS 2010, Italy, Jun. 2010.