

BitTorrent Darknets

Chao Zhang*, Prithula Dhungel*, Di Wu†, Zhengye Liu* and Keith W. Ross*

* Polytechnic Institute of NYU, Brooklyn, NY

† Sun Yat-Sen University, Guangzhou, China

Abstract—A private BitTorrent site (also known as a “BitTorrent darknet”) is a collection of torrents that can only be accessed by members of the darknet community. The private BitTorrent sites also have incentive policies which encourage users to continue to seed files after completing downloading. Although there are at least 800 independent BitTorrent darknets in the Internet, they have received little attention in the research community to date. We examine BitTorrent darknets from macroscopic, medium-scopic and microscopic perspectives. For the macroscopic analysis, we consider 800+ private sites to obtain a broad picture of the darknet landscape, and obtain a rough estimate of the total number of files, accounts, and simultaneous peers within the entire darknet landscape. Although the size of each private site is relatively small, we find the aggregate size of the darknet landscape to be surprisingly large. For the medium-scopic analysis, we investigate content overlap between four private sites and the public BitTorrent ecosystem. For the microscopic analysis, we explore in-depth one private site and examine its user behavior. We observe that the seed-to-leecher ratios and upload-to-download ratios are much higher than in the public ecosystem. The macroscopic, medium-scopic and microscopic analyses when combined provide a vivid picture of the darknet landscape, and provide insight into how the darknet landscape differs from the public BitTorrent ecosystem.

I. INTRODUCTION

BitTorrent is a remarkably popular file-distribution technology, with millions of users sharing content in hundreds of thousands of torrents on a daily basis. Even in the era of YouTube, BitTorrent traffic continues to grow at impressive rates. For example, downloads of .torrent files from Mininova’s site doubled in 2008, to nearly 7 million downloads in a year [1]. BitTorrent has proven to be particularly effective at distributing large files, including open-source software distributions. In an earlier paper, we presented results from a comprehensive crawl of public BitTorrent Ecosystem [2]. In that study, by crawling five of the most popular Torrent discovery sites (namely, Pirate Bay, Mininova, Torrent Reactor, BTmonster, and Torrent Portal), we discovered 4.6 unique torrents and 39,000 unique trackers. Figure 1 shows the components of the BitTorrent ecosystem.

In this paper we seek to obtain an in-depth understanding of BitTorrent private sites (also known as “BitTorrent darknets”). A BitTorrent private site restricts who can use it, typically by requiring users to register accounts. Many of these sites use invitation systems for limiting registrations. Private torrent sites typically record how much the registered users upload and download, and enforce a minimum upload-to-download ratio on each user. These ratio policies provide incentives to users to continue to seed files after downloading. Although there are hundreds of independent BitTorrent darknets in the

Internet today, they have received little attention in the research community to date – little is known about their scope and behavior within the Internet.

To our knowledge, this is the first measurement study of BitTorrent darknets. Such a study is important, as a big-picture understanding of the BitTorrent Ecosystem must take into account both the public and private worlds within BitTorrent. For the macroscopic analysis, we consider 800+ private sites to obtain a broad picture of the darknet landscape, including geographic concentrations and content distributions. We perform a regression analysis that correlates a site’s Alexa rank with its number of files shared, number of registered users, and number of active peers. This allows us to obtain a rough estimate of the total number of files, accounts, and simultaneous peers across the darknet landscape. For the medium-scopic analysis, we perform a measurement study of four popular private BitTorrent websites, identifying the characteristics of over 700,000 torrent files that these websites host. For each of these four darknets, we also crawl their trackers, identifying the set of peers participating in each torrent. Using the data set obtained by crawling the webpages and torrents, we analyze the trackers, peers, users, and content that the BitTorrent darknets embody. By also measuring the torrents and peers in the public BitTorrent, we further compare and correlate both the public and private worlds. For the microscopic analysis, we explore in-depth one private site and examine how the incentive policy influences user behavior. The macroscopic, medium-scopic and micro-scopic analyses not only present a vivid picture of the darknet landscape, but also provide insight into how the BitTorrent darknets differ from their public counterparts.

This paper is organized as follows. Section II provides an overview of BitTorrent darknet operation. Section III presents a macroscopic study of BitTorrent darknets. In Section IV, we conduct a medium-scopic measurement study of four popular private sites. In Section V, we examine a typical private site - HDChina - from a microscopic view. In Section VI, we review prior work, and conclude in Section VII.

II. OVERVIEW OF BITTORRENT DARKNET OPERATION

Typically the “owner” of a BitTorrent darknet operates and manages both a torrent-discovery site (Web site) and a tracker. For users to browse the darknet’s torrent-discovery site and to obtain .torrent files indexed on the site, they must first register with the site with a login name and password. For many of the darknet sites, a user first needs to obtain an invitation before it can register at the site. After registration, and once the user

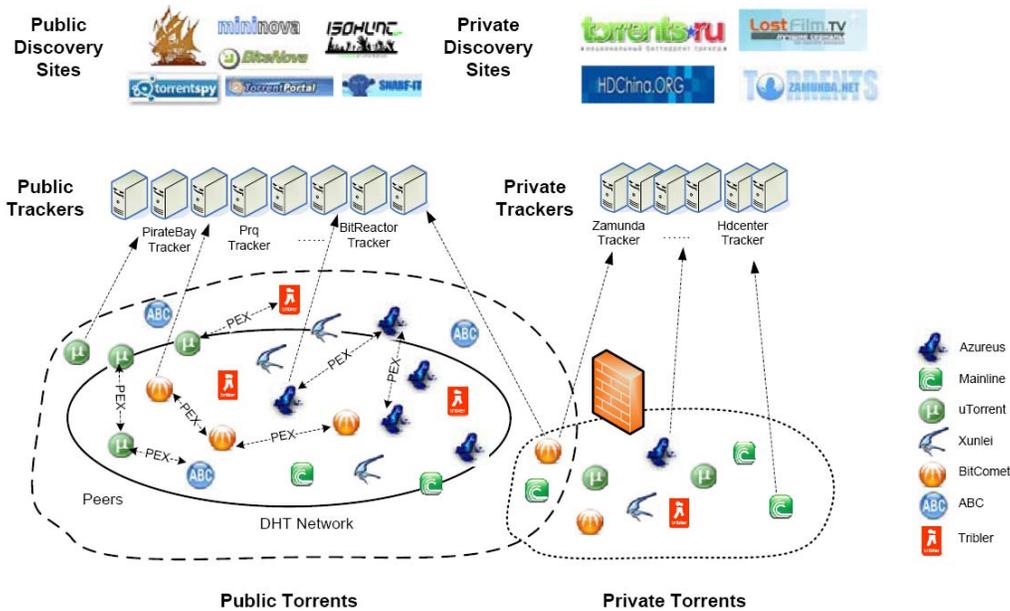


Fig. 1. The BitTorrent Ecosystem.

proves itself as a good citizen within the darknet, typically by uploading many files, the user in turn may be provided with invitations, which it can distribute to friends (or sometimes sell over eBay!). This procedure is very different from a public discovery site (such as Mininova or PirateBay), which allows arbitrary users to browse the site and download .torrent files without registration.

The private sites assign each registered user a unique “passkey”. This passkey provides authorization to the darknet’s tracker. As an example, suppose Joe is a registered user at the fictional darknet DVDdarknet.com, and is interested in obtaining the movie FantasticMovie. Joe will first browse DVDdarknet, and request the metadata .torrent file for FantasticMovie. The .torrent file returned by the DVDdarknet.com will contain the IP address (or hostname) of the private tracker, which is also managed by DVDdarknet, as well as other meta-information. In particular, the .torrent file will include Joe’s passkey and will typically have the “private field” set to one. Joe’s BitTorrent client then reads the .torrent file, contacts the private tracker provided in the file, and provides Joe’s passkey for authorization. The tracker, after having validated the passkey, provides Joe with a subset of peers (typically up to 50) that are currently active in the swarm for FantasticMovie. Joe’s BitTorrent client then establishes TCP connections and trades blocks of data with these other peers, which are also registered at DVDdarknet.com.

When joining a swarm in the public BitTorrent Ecosystem, many popular BitTorrent clients will also join a DHT (Distributed Hash Table), registering itself in the DHT as a participant in the swarm. It will also use the PEX (Peer Exchange) protocol to gossip with other peers in the swarm about the other active peers in the swarm that it knows about. This is very different from a BitTorrent darknet, for which the private flag indicates to the BitTorrent client that it should not join a DHT or gossip with PEX. However, we shall see that

this rule is not always respected by all darknets, and that peers in private torrents sometimes leak their IP addresses and port numbers to the outside world via DHTs and PEX.

Typically a private site does not develop itself the tracker and torrent-discovery software, but instead installs one of the many open-source implementations currently available [3]. For example, the XBTT tracker [4] has many desirable features, such as integrated bulletin board system and one click installation, which enable site administrators without any coding skills to quickly and easily run their own BitTorrent trackers.

One major challenge for exploring a darknet is getting a membership account. Many of these sites limit the number of user accounts, only making available new accounts when existing accounts are terminated. There exist some tools (e.g., Tracker Checker[5]) and web sites (e.g., BTRACS [6]) that can notify users when new membership accounts open up. Some darknets require user invitations, which makes it even harder to get accounts. Some sites further prevent multiple accounts from the same IP address, hindering aggressive crawling from a single IP address.

A. Incentive Policies of Private Sites

A distinctive characteristic of a BitTorrent darknet is the “ratio incentive” policy. Specifically, when a user participates in a swarm managed by a darknet, its client periodically reports to the darknet tracker, along with its passkey, the amount of data it has uploaded and downloaded from the swarm. The darknet uses this information to track the total number of bytes the user has uploaded and downloaded for *all* the torrents (from the darknet) the user has participated in to date. For each of its registered users, the darknet maintains the ratio of the number of bytes it has uploaded to the number of bytes the user has downloaded. Users are given incentives to maintain a high ratio. Most sites provide high-ratio users with elevated

treatment such as the ability to download the latest torrents without having to wait and enhanced browsing capabilities in the private site. Furthermore, these sites also mandate the users to maintain a minimum ratio value (e.g., 0.9) in order to keep their accounts active. This incentive mechanism not only encourages darknet users to seed previously-downloaded content into the darknet, but also to locate and seed fresh content that will appeal to other darknet members. We stress that this ratio incentive policy is unique to private torrents – the public sites (PirateBay, Mininova, and so on) do not record users’ upload and download contributions across torrents and do not enforce a minimum ratio.

III. MACROSCOPIC ANALYSIS

In this section, we seek to obtain a big-picture understanding of the scale of the BitTorrent darknet world. Ideally, we would like to get a rough idea of how many BitTorrent darknets there are, how many files are being shared across the totality of darknets, how many users participate across the totality of darknets, where the trackers of the darknets are located, where the users of the darknets are located, and so on.

This is a very challenging problem for many reasons. First, it is very hard to identify all the darknet sites, as many of them attempt to be as inconspicuous as possible. Any list of darknets will necessarily be incomplete. Second, to obtain meaningful information about a darknet, we need to become a member of its site. As described in Section 2, gaining membership can only be done manually and may be impossible without an invitation. Third, even if we gain membership to the darknet, the statistics reported in the darknet about the number of torrents, registered users, or active users may be incomplete or inaccurate. Nevertheless, in this section we present a methodology which gives a rough, ball-park estimate of the scale of the BitTorrent darknet world.

To estimate the scale of the BitTorrent darknet world, we begin with a recent list of 900+ sites compiled by Sharky [7]. As described in [8], Sharky laboriously and carefully compiled the list in June 2009. To create the list, Sharky used a variety of different information sources, including tracker checker websites, file sharing blogs and forums, Google searches using carefully crafted search operators, invite-sharing queue sites, tracker monitoring software utilities, and IRC invite channels. Using such information sources, Sharky maintains a list of private trackers, and periodically updates it by manually checking if the existing sites function and adding newly discovered site. Sharky also classifies each site into one of fourteen content categories.

Sharky’s list is the starting point for our macroscopic analysis. We emphasize, however, that this list is to some degree incomplete, as there are certainly many stealth private torrents that do not advertise their presence in any public venue. Nevertheless, Sharky’s list most likely contains all the largest private torrents, since large darknets are eventually noted in one of the many public venues that Sharky tracks. We preprocessed Sharky’s list, manually checking each site and keeping those that are only operational in July 2009, creating

Category	# of sites	Category	# of sites
General	473	TV	28
Application	15	Misc	18
Anime	21	Movie	35
Desi	15	Music	88
Elearning	23	Scene	17
Games	19	Sports	31
Porn	47	HD	24

TABLE I
PRIVATE SITE SPECIALIZATION

a processed list of 863 unique private sites. Table I shows the number of private sites in each category. Note that about 55% of the private sites contain general content; the remaining 45% of the private sites contain specialized content.

Using the MaxMind GeoIP tool [9], we mapped each host on the list to its geographic location. Table II shows the number of private sites in each country. We observe that a significant fraction of the private sites are located in Europe, with Netherlands leading the way, by far, for the number of sites per capita.

Country	# of sites	Country	# of sites
USA	194	Malaysia	15
Netherlands	107	Luxembourg	14
Germany	83	Ukraine	10
Sweden	58	Thailand	9
France	50	Bulgaria	8
Canada	46	Denmark	8
Hungary	44	China	8
Romania	36	Czech Republic	7
UK	36	Slovenia	7
Russia	16	Others	117

TABLE II
GEOGRAPHIC DISTRIBUTION OF PRIVATE SITES

The next step in our methodology is to use Alexa [10] to gain insight into the rough popularity of each of the private sites on the processed list. Alexa presents usage statistics on Web sites in the Internet. To obtain this information, Alexa publishes a tool that can be integrated to Web browsers. After an Internet user installs the tool, the tool determines which sites the user has visited, as well as the number of visits for each site, and reports this information to the Alexa server. Alexa then ranks the Web sites based on the information gathered from all users that have installed the Alexa tool. We developed an automatic scraping script to retrieve Alexa information for each private torrent site. We found that Alexa may mis-report traffic information for some sites (for example, it sometimes reports a superdomain’s traffic rather than for the domain itself). We removed the sites that appear to have mis-reported information, and finally have a list of 797 sites.

Table III presents the 15 most popular sites based on the Alexa rank. It is interesting to observe that six of these sites are based in the Netherlands. We also observe that two of the most popular sites specialize in pornography. Surprisingly, only one Chinese site is in the top fifteen. This may be because digital content is largely available from Xunlei [11] and other sources in China. Although we were not able to register at each of the 15 sites, we were still able to obtain information from their welcoming Web pages and from BitTorrent user forums. The Russian-language site Torrents.ru appears to be

Site	Language	Alexa rank	Site Location	Theme
Torrents.ru	Russian	295	Russia	General
Zamunda.net	Bulgarian	1,760	Netherlands	General
PureTnA.com	English	2,291	Netherlands	Porn
gamato.info	English	2,751	Netherlands	General
Emporium.us	English	2,846	Netherlands	Porn
arenabg.com	Bulgarian	4,267	Netherlands	General
lostfilm.tv	Russian	4,593	Russia	TV, Movies
bwtorrents.com	English	5,414	USA	General
www.gamato.info	Greek	5,493	Netherlands	General
torrentleech.org	English	5,989	USA	General
Bitsoup	English	6,762	Canada	General
www.h33t.com	English	7,663	Netherlands	General
bittorrents.ro	Romanian	9,234	USA	General
HDChina	Chinese	14,305	China	HD Video

TABLE III
THE 15 MOST POPULAR SITES BASED ON ALEXA RANK

the largest BitTorrent darknet in existence today. It indexes a broad range of content, including movies, television shows, music, games and so on. Much of this content is copyright protected. By obtaining an account on the site, we discovered that it has over 612,000 torrents and about 3.5 million user accounts.

Country	Percentage of users	Country	Percentage of users
Russia	30.7%	Bulgaria	3.8%
USA	17.4%	Greece	2.9%
India	8.3%	UK	2.6%
Ukraine	4.9%	Turkey	2.3%
Japan	4.1%	Romania	2.2%

TABLE IV
DARKNET USAGE BY COUNTRY

Alexa also provides information about the countries from which the sites's visitors come. Combining this Alexa data with the processed list, enables us to estimate what countries are using private torrents the most. Table IV provides the percentage of users from each of top-10 private-torrent usage countries. We observe that Russia leads the way in BitTorrent darknet usage. Interestingly, the Netherlands isn't a top-10 usage country, even though it has the second largest number of sites (including many of the largest sites).

A. Regression Analysis and Totality Estimates

We next explore whether there is a statistical correlation between the Alexa rank and the size of the site (that is, the number of torrents it indexes, the number of active users, and so on). To this end, we randomly selected a set of 67 sites from the list and attempted to obtain accounts for each of the sites. We obtained accounts for 33 sites. For the sites that we were able to join, we manually browsed the site's statistics (often using Google to translate the site to English) to obtain the number of registered accounts, the number of torrents, and the number of active peers. Some of the sites only provided partial statistics. Using this sample data, we conducted a regression analysis using Matlab toolboxes to analyze the correlation between the Alexa traffic rank and private torrent site statistics.

Figure 2 shows the results obtained from our regression analysis. Note that for three scale measures – number of torrents, number of registered users, and number of active peers – the R^2 value ranges from 0.81 to 0.89, indicating a

strong correlation between the Alexa rank and the three scale measures. Moreover, we see that distributions for each measure is Zipf-like, implying that the site popularity distribution across the 797 sites has a thick tail. (For the number of active peers, a given peer may be counted multiple times at a site, once for each torrent it belongs to.)

Let x be the Alexa rank of a given private torrent site, y_t be the number of torrents of that site, y_a be the number of registered accounts of that site, and y_p be the number of active peers of that site. Using the regression analysis, we obtain the following correlation equations:

$$\begin{cases} y_t = 10^{7.17} \cdot x^{-0.695} \\ y_a = 10^{8.56} \cdot x^{-0.871} \\ y_p = 10^{9.75} \cdot x^{-1.173} \end{cases} \quad (1)$$

Given the Alexa rank, and using the above equations, we can estimate the number of torrents, registered accounts and peers in each of the private torrent sites on our processed list. We can then get a rough estimate of the total number of torrents, registered accounts, and peers in BitTorrent darknets by aggregating the estimation results in the private sites. Table V shows the estimate of the total number of torrents, registered accounts, and peers in the BitTorrent darknets. Although each individual private site is small, the aggregate number of torrents (or registered accounts, active peers) is remarkably large. According to this rough estimate, there are about 4.4 million torrents, 20 million registered accounts, and over 24 million active peers in the BitTorrent darknets. As the users of private sites prefer to seed a file after downloading and thus join multiple torrents simultaneously, the number of active peers is higher than the number of registered accounts. Of course, this estimate also double-counts across sites. For example, a peer may be active in multiple private sites simultaneously, or a user may have accounts on multiple private sites. And, as mentioned earlier, the list is certainly not complete, as it only includes private sites that have been noted in a public Internet venue. Nevertheless, these numbers provide a rough, ball-park estimate of the scale of private torrent landscape.

	Estimate
Total # of torrents	4,413,719
Total # of registered accounts	20,451,849
Total # of active peers	24,716,652

TABLE V
ESTIMATE OF THE TOTAL NUMBER OF TORRENTS, REGISTERED ACCOUNTS, AND PEERS IN THE BITTORRENT DARKNETS

In a separate paper, we performed a comprehensive study of the public BitTorrent landscape[2]. In that study we crawled five top public torrent sites: *Mininova*, *Pirate Bay*, *Torrent Reactor*, *BTmonster*, and *Torrent Portal*. They are, respectively, the first, second, fifth, seventh, and ninth most popular English-language public torrent sites. We collected approximately 8.8 million .torrent files from these public torrent sites, from which we obtained 4.6 million unique infohashes and 38,996 trackers. We also obtained a snapshot of each active torrent over a twelve-hour period. In one typical snapshot, the total number

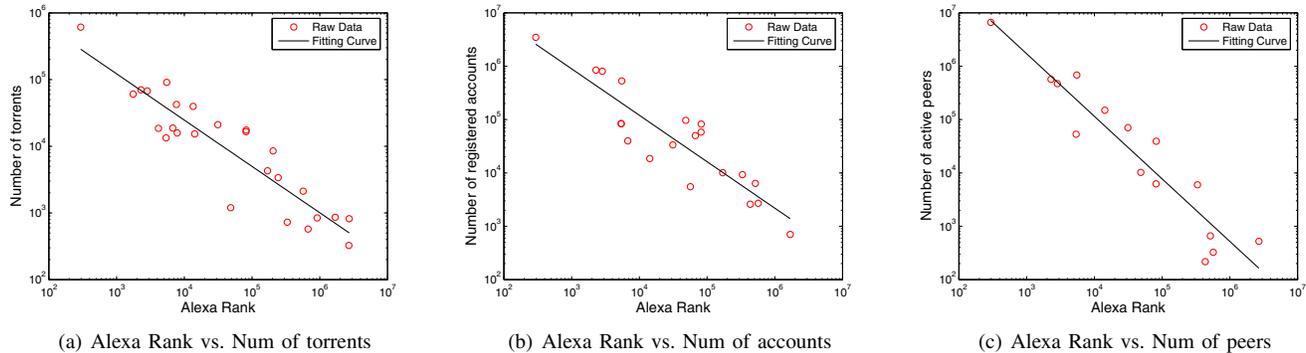


Fig. 2. Regression Analysis: (a) Alexa rank vs. number of torrents (Goodness of fit: $R^2 = 0.84$); (b) Alexa rank vs. number of registered accounts (Goodness of fit: $R^2 = 0.81$); (c) Alexa rank vs. number of active peers (Goodness of fit: $R^2 = 0.89$);

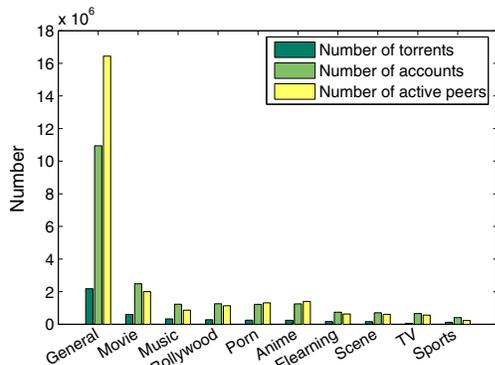


Fig. 3. Estimation of the number of torrents/accounts/active peers in different categories

of unique peers observed was 5,085,217. From our rough regression-based estimate, we see that the number of torrents hosted in the BitTorrent private world is comparable to that in the public (English-language) world. Moreover, the number of active peers in the private world is much larger than that in the public English-language world!

Using the content theme information that we have for each of the private torrents, we can further estimate the number of torrents, registered accounts, and active peers in different content categories. Figure 3 shows these estimates.

In summary, we have presented a methodology that combines (i) collecting private site URLs made available from a range of Internet venues (which Sharky has done for us), (ii) Alexa web site ranking, and (iii) regression analysis to obtain a high-level, informative picture of the darknet landscape, consisting of more than 800 private sites.

IV. MEDIUM-SCOPIC ANALYSIS

In this section, we conduct medium-scopic measurements for BitTorrent darknets by crawling the torrent-discovery sites and trackers for a small number of private sites.

A. Challenges in Web Site and Tracker Crawling

There are many challenges in crawling a darknet’s torrent-discovery site (Web site) and downloading the .torrent files it indexes. Some darknet torrent-discovery sites limit the maximum number of .torrent files a member can download in a certain time (e.g., at most one .torrent per minute from a single account). Possessing a small number of accounts

(due to the challenges of account registration as described above) further exacerbates this problem since, with just a few accounts in hand, the maximum crawling speed is often slow. Furthermore, each of the private sites is organized differently, with important meta-data (e.g., uploader name, date of upload, category, and so on) displayed in different manners. This heterogeneity among the darknet Web site organization and layouts requires us to develop separate parsers for each site investigated. Furthermore, many of the darknet Web sites are presented in languages other than English, which further complicates the parsing of the Web pages.

B. Medium-scopic sites

We crawled four popular private torrent-discovery Web sites: Torrents.ru, Zamunda, BitSoup, HDChina. We chose these sites because (1) they are all among the top 15 private-torrent sites according to Alexa; (2) we were able to get accounts on these sites; (3) the sites have a range of geographical locations and language bases; (4) the sites have a range of content specialization. For each of these sites, we downloaded all of the .torrent files it indexed.

From April 11, 2009 to June 13, 2009, we also crawled the trackers of Zamunda, BitSoup, HDChina (that is, repeatedly asked for peer lists). Each of these private sites uses one tracker. Because crawling the Torrents.ru tracker is very slow, we took an entirely different approach for it. Unlike most darknets, Torrents.ru does not set the private flag to 1 in the .torrent file. Therefore, the BitTorrent clients register with their DHTs for Torrent.ru swarms. To get the IP addresses of the peers in the Torrents.ru swarms, we crawl the DHT instead of the tracker. For each of these four darknets, Table VI shows the number of torrents it indexes, the Web site’s language, and its theme. We remark that even though the Web sites for Torrents.ru, Zamunda, HDChina are not in English, the actual audio in the content for most of their torrents is in English. We define an *active torrent* to be a torrent that has at least one active peer.

C. Overlap and Leakage with the Public Ecosystem

The first question we pose at the medium-scopic level is: Are many of the files being distributed in the darknets also distributed in the public Ecosystem, as engendered by the top-5

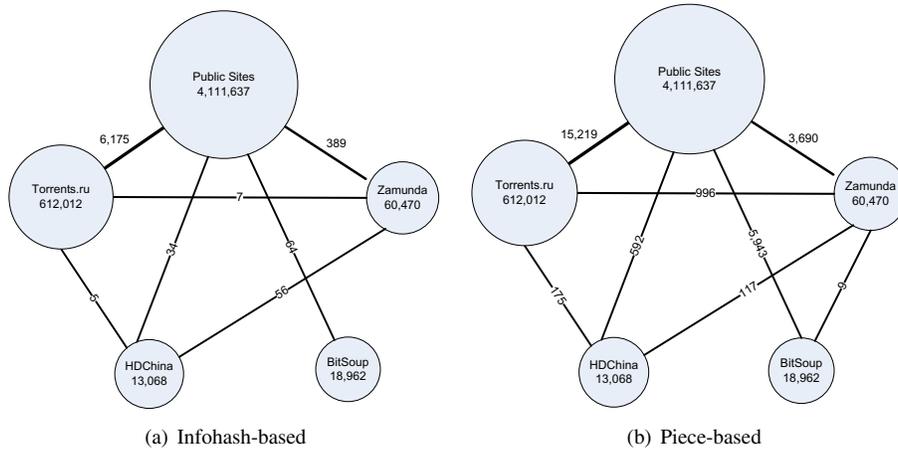


Fig. 4. Pairwise-intersection of active torrents for private and public sites: (a) Infohash-based; (b) Piece-based

Site	# of torrents	# of active torrents	Language	Theme
Torrents.ru	612,012	294,201	Russian	General
Zamunda	60,470	32,120	Bulgarian	General
BitSoup	18,770	6,937	English	General
HDChina	13,068	6,459	Chinese	HD video

TABLE VI

SUMMARY OF DARKNETS FOR MEDIUM-SCOPIC ANALYSIS

public torrent-discovery sites? Figure 4(a) shows the pairwise-intersection of files based on their infohashes for the four private sites and for the public Ecosystem. Because a private-flag value of 1 is sometimes included in the calculation, a simple infohash comparison may not identify all the files that are both in a private site and the public ecosystem. For this reason, we also identify matches using the set of hashes of the individual pieces of the file, and provide the results in as Figure 4(b). Note in Figure 4(b), the absence of a link mean that there is no overlap between the two sites.

We make the following observations. First, the hashes of the individual pieces identifies more matches than do the infohashes. Therefore, in analyzing file overlap, it is better to use hashes of individual pieces. Second, there is very little file overlap among the four private sites. This finding is completely different from what we found in the public BitTorrent ecosystem [2], where the intersection between any two public sites was more than 50%. The homogeneity in the public ecosystem is primarily due to two factors: (i) users often upload the same .torrent file to multiple torrent-discovery sites; (ii) some torrent discovery sites crawl other torrent discovery sites, obtain their .torrent files, and index them on their own sites. In darknets, due to differences in content and language themes among them, there is little overlap of the exact files (although there is more significant overlap of content, as we will demonstrate shortly). Third, the intersection between private and public sites is relatively small compared to the total number of files provided by each private site. For example, only one third of the files in BitSoup, an English-language site, are available in the public Ecosystem. This observation shows that although many users active in both the public ecosystem and in darknets, there is little direct “copying” from the public to private worlds and vice versa.

The file overlap does not present the complete picture since it is possible for two files to have a different set of hashes even though the two files are essentially the same. For example, two sites may distribute the same DVD, but with different regional versions using a different language for menus, subtitles and so on. To obtain a more complete understanding of the content overlap among the sites, we also check the matching of torrent names and their sizes (in bytes). An example torrent name is “Ghost Ship. HDDVD.1080p.DTS.x264-CtrlHD”. Note that a typical torrent name contains not only the title of the content (“Ghost Ship” in this example), but also the codec, resolution and so on. We say that two torrents indexed on different sites have a *title match* if the torrent names use the same title (e.g., both use “Ghost Ship”). We say that two files have an *extended match* if they have not only the same title but also the same file size (within 5%) and (if presented in the torrent name) the same codec and resolution. For many reasons, two files may have the same extended match but not have the same hash sets; for example, if the two files were derived from two different regional versions of the same DVD. Similarly, two files may title match but not extended match; for example, the two files contain the same movie but encoded at very different resolutions.

To investigate title matching and extended matching among the sites, we select the top-100 and a random 100 torrents from Zamunda, BitSoup, and HDChina, which all use torrent names (including titles) in English, and check the intersection with the torrents in the public Ecosystem. Torrents.ru is ignored because the torrent names are given in Russian. (We tried translating the Russian titles. But the translations led to inaccuracies and therefore inconclusive results.) To determine whether the torrent is available in the public ecosystem, we input the title into the comprehensive torrentz discovery site and then manually examine the results.

Table VII presents the title matching and extended matching results. We see that for the top-100 files on each of the three private sites, versions of most of these files indeed exist in the public ecosystem. Specifically, the title matching is 96% or higher for all three sites. For the random 100 files, title matching remains high for HDChina, but drops to 64% and

76% for Zamunda and BitSoup, respectively. One reason for the drop off in Zamunda and BitSoup is that older TV episodes of the same TV series may be combined and renamed in different ways. So the actual content overlap for random 100 is actually higher for these two sites. HDChina does not have this recombination issue, as it mostly distributes HD and Blu-ray movie content. In summary, with respect to title matches, most of the titles that are available in the three private sites are also available in the public ecosystem.

Table VII shows that Zamunda and BitSoup have extended matches with the public ecosystem that are just a little less than the corresponding title matches (for top-100 and random). This implies that for each of these two private sites, essentially the same versions of the content are available in the private site and the public ecosystem. However, for HDChina there is a significant difference between title matching and extended matching. This is likely because for many of HDChina’s movies, they are distributed in high-resolution (DVD and Blu-ray) in HDChina and lower resolution in the public ecosystem. Because HDChina torrents have a large percentage of seeds (due the incentive mechanism, as discussed in Section V), it is possible for users to download the gigantic HD and Blu-ray files much faster in HDChina than in the public ecosystem.

Darknet	Top-100		Random	
	TM	EM	TM	EM
Zamunda	97	94	64	53
BitSoup	96	92	76	71
HDChina	97	65	91	34

TABLE VII

INTERSECTION OF ACTIVE TORRENTS FOR PRIVATE AND PUBLIC ECOSYSTEM: TITLE MATCHING (TM); EXTENDED MATCHING (EM)

The next question we pose is: Is it possible to discover darknet peers and obtain content from them without registering at the corresponding darknet sites? One approach for bypassing the darknet registration is to download from private peers whose IPs have leaked into a DHT. To evaluate this leakage, we developed a DHT crawler to crawl the DHT system for all the infohashes obtained from private sites. The leakage status is reported in Table VIII. We observe that Torrents.ru has a 100% leakage rate into the DHT. This is because Torrents.ru does not set the private flag in its .torrent files. We also observe that all the other private sites have a low rate of leakage.

Site	# of torrents observed in DHT	Fraction of leaked torrents	# of peers observed in DHT
Torrents.ru	294,201	100%	2,030,583
Zamunda	361	1.11%	820
Bitsoup	344	4.95%	643
HDChina	307	4.75%	5247

TABLE VIII

LEAKAGE OF PRIVATE SITES IN DHT (NOTE: IN THIS TABLE, A PEER REFERS TO A TRIPLE <INFOHASH, IP, PORT>.)

In summary, we have found that most of the content being distributed in the private sites (at least for the four sites investigated in this paper) is also present in the public ecosystem. However, demonstrated by the low overlap rates for the hashes, the versions in the public ecosystem are

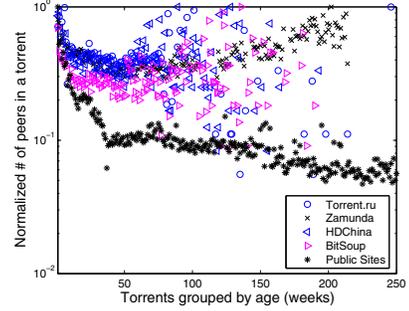


Fig. 5. Active torrent size (normalized)

typically different from those in the private ecosystem. For some private sites, the versions only differ from what is available in the public ecosystem in minor ways (for example, different regional distributions of the same DVD). For other private sites (such as HDChina), the versions may have more significant differences, for example, major difference in video resolution. Also, although direct copying from public to private worlds (and vice versa) does occur, it appears to be relatively infrequent.

D. Characteristics of Private Torrents

We now examine the popularity of private torrents as a function of their lifetimes. To this end, we consider torrents to be in the same group if their .torrent files were uploaded in the same week. Figure 5 shows the average torrent size per group for private sites and public sites. So that we can compare each of the sites more easily, we have normalized each value by its maximum value. Figure 5 shows that, in both the public ecosystem and the private sites, the newly released torrents, on average, attract more peers. Although older torrents attract fewer peers, the decay in popularity of private torrents is much less dramatic than with in public torrents. This can likely be attributed to the purging policies on private sites – administrators of private sites may, at their discretion, remove unpopular torrents, which increases the average popularity of the remaining torrents. Figure 6 shows the CDF of torrents in different ages. The average torrent age on private sites is smaller than that on public sites. This is again likely because private site administrators often purge older, less popular torrents. The older, less popular torrents can persist for years in the public Ecosystem. Figure 7 shows the number of peers in a torrent, ordered from the largest to smallest one. Note that all the sites follow Zipf-like distribution, and that private sites have a longer heavy tail.

Figure 8 shows the distribution of BitTorrent client types used to create .torrent files for private sites. Compared with the usage of different client software in public sites, we observe that uTorrent is even more widely used in the darknets, with about 82% of the torrent creators being uTorrent users (as compared with 64% in public sites). The usage of Azureus (called Vuze now) drops significantly from 16% to 4%. A possible reason for these changes is private site users, who tend to be more power BitTorrent users, prefer to use the lighter uTorrent client to the Java-based, feature-rich Azureus client, which consumes more CPU and memory resource. We also

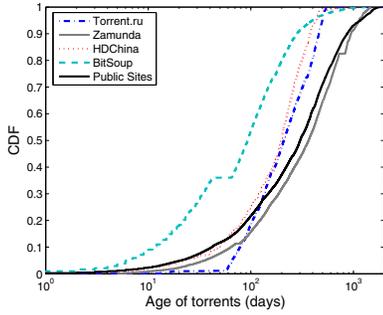


Fig. 6. Lifetime distribution of torrents on different sites

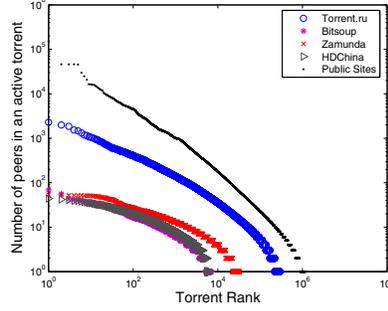


Fig. 7. Ranking of active torrents on private and public sites

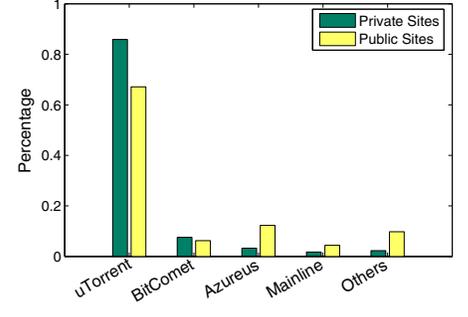


Fig. 8. Distribution of BitTorrent client types used to create .torrent files for private and public sites

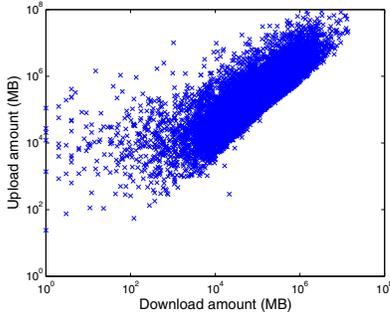


Fig. 9. Download/upload amount of peers (MB)

remark that many private sites require their members to use (closed-source) uTorrent.

V. MICROSCOPIC ANALYSIS

In this section we complement our macroscopic and medium-scopic analyses with a detailed analysis of one private site, namely, HDChina. HDChina is a large private torrent site that mainly distributes high-definition movies and TV episodes. At the time of our measurement, HDChina had 18,504 registered accounts and 15,738 active torrents. In order to join HDChina, the user must obtain an invitation from another senior member of HDChina. Similar to other private sites, HDChina also implements a ratio incentive policy. For users whose download amount is less than 10GB, they need to maintain a ratio (upload amount/download amount) higher than 0.3. The required ratio increases when the user downloads more content: if the user has downloaded more than 100GB, the required ratio is as high as 0.7. If the user fails to satisfy the required ratio, he/she will receive a warning email from the administrator. If the ratio is still low one week after being warned, the user will be banned by the system.

HDChina makes available on its site extensive information about individual user usage. For HDChina, we additionally crawled and parsed the user data for all of the HDChina users. With this additional user data, we can provide a more detailed analysis of the characteristics of BitTorrent darknets and examine closely the consequences of its ratio incentive policy. Figure 9 provides a scatter plot, showing for each registered user the amount it downloaded and uploaded. The upload amount of most users is higher than their download amount. This can be attributed to the ratio incentive policy. In fact, a large fraction of users have uploaded 50+ MBytes more

than they have downloaded. These users appear to be building up credit, which can be used for future bursts of desirable content. Remarkably, many users have uploaded more than 1 TB data.

In HDChina, the aggregate amount of uploaded bytes and the aggregate amount of downloaded bytes are 17,054 TB and 2,568 TB, respectively. In principle, the aggregate amount of data uploaded should equal to the aggregate amount downloaded. But the aggregate amount of uploaded bytes greatly exceeds the aggregate amount of downloaded bytes. There are multiple possible reasons for this: (1) newcomers are allowed to download their first file for free, which is not included into the download amount; (2) users who download much more than they upload get their accounts removed by the administrator but nevertheless contribute to aggregate download and upload measures.

Figure 10 shows the CDF of the share ratio of all the registered users. Over 90% of users have a ratio higher than 1, and less than 5% of users have a ratio higher than 100. Figure 11 shows the CDF of the last online time of registered users. Observe that about 50% of users are online within 10 hours of our crawling, and 95% of users are online within 100 hours of our crawling. Thus, the ratio incentive policy is causing users to be very active.

Figure 12 shows the rank of uploaders on PirateBay and HDChina based on the number of uploaded torrents, which exhibits clear Zipf-like distribution. Although the top uploaders in HDChina upload fewer torrents than that in PirateBay, the Zipf parameter of HDChina is much bigger that of PirateBay.

In summary, the microscopic measurement of HDChina enables us to gain new insights into user behavior in private sites. We observe that the ratio policy is very effective in incentivizing user uploading, and over 90% of users have a ratio higher than 1. In a companion paper [12], we examine this site in greater detail and, in particular, show that the seed/leecher ratio in private sites is significantly higher than in public sites, thereby providing private sites with accelerated download speeds.

VI. RELATED WORK

In recent years, a number of studies have been conducted measuring different aspects of the BitTorrent ecosystem. Falkner et al. [13] performed a measurement study of the Azureus DHT and analyzed peer churn, overhead, and per-

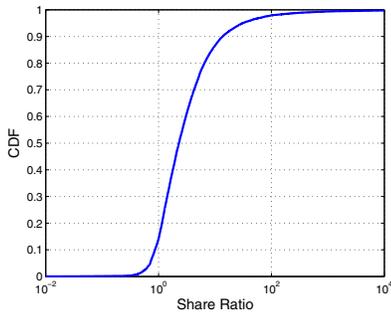


Fig. 10. Distribution of share ratio among peers

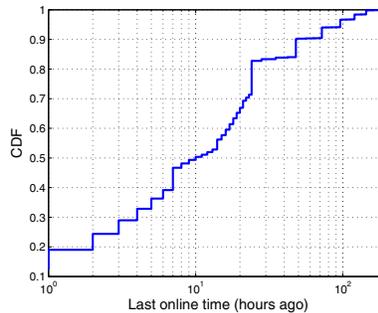


Fig. 11. CDF of last online time (hours ago)

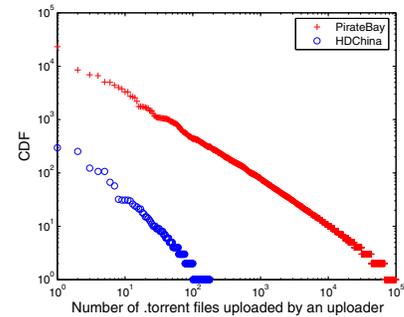


Fig. 12. Uploader rank based on the number of uploaded torrents

formance in the DHT. The authors of [14] looked at the two BitTorrent DHTs – Mainline and Azureus – and analyzed peer churn, latency, and liveness of nodes. There have also been several papers on measurements involving public trackers. Pouwelse et al. study BitTorrent usage, using statistics gathered from a single torrent-search site (the now defunct Suprnova site) [15]. Bellissimo et al. collected 3-month logs using data gathered from two trackers and investigated a limited number of torrent and peer characteristics [16]. Guo et al. [17], [18] measured torrent evolution, service availability, and client performance by analyzing a limited number of tracker traces from [16] and torrent file download traces. Neglia et al. [19] investigated the availability of BitTorrent systems; they collected about 22,000 torrents from two torrent-discovery sites and mainly focused on tracker/DHT reliability issues. Izal et al. analyzed the behavior of a single torrent over a five-month period [20]. In an earlier paper [2], we performed a large-scale measurement study to provide a near-complete picture of the public BitTorrent ecosystem, studying in-depth the ecosystem’s torrent-discovery, tracker, peer, user behavior, and content landscapes.

However, to best of our knowledge, there has been no work to date on the measurement and analysis of the more covert side of BitTorrent – the private BitTorrent world. More recently, we have used mechanism design to study incentive policies in darknets and have also studied collusion detection mechanisms [12].

VII. CONCLUSION

In this paper, we conducted a comprehensive measurement study of BitTorrent darknets from macroscopic, medium-scopic and microscopic perspectives. In our macroscopic analysis, we investigated 800+ private sites in terms of their geographic concentrations and content distributions. We presented a methodology that combines collecting private site URLs made available from a range of Internet venues, Alexa web site ranking, and regression analysis to obtain a high-level, informative picture of the darknet landscape, consisting of more than 800 private sites. Although the size of each private site is relatively small, we find the aggregate size of the darknets is surprisingly large. In our medium-scopic analysis, we crawled the web sites and trackers of four popular private sites, and compared their statistics with those of the public sites. We also investigated content overlap among these

sites and also the overlap between four private sites and the public BitTorrent ecosystem. In our microscopic analysis, we examined how the ratio incentive policy influences user behavior. We observe that the seed-to-leecher ratios and upload-to-download ratios are much higher in the private sites than in the public ecosystem. Combining macroscopic, medium-scopic and microscopic analysis together, we have presented a clear and comprehensive picture of the BitTorrent darknet landscape.

REFERENCES

- [1] “Mininova’s Torrent Downloads Double to 7 Billion in a Year,” <http://torrentfreak.com/mininovas-torrent-downloads-doubled-in-a-year-090105>.
- [2] C. Zhang, P. Dhungel, D. Wu, and K. W. Ross, “Unraveling the BitTorrent Ecosystem,” in *Technical Report, Polytechnic Institute of NYU*, May 2009, <http://cis.poly.edu/~chao/bt-ecosys/bt-ecosyste-TR.pdf>.
- [3] “Bittorrent tracker software,” http://en.wikipedia.org/wiki/Comparison_of_BitTorrent_tracker_software.
- [4] “XBTTTracker,” <http://xbtt.sourceforge.net/tracker/>.
- [5] “Tracker Checker,” <http://www.brothersoft.com/tracker-checker-64477.html>.
- [6] “BTRACS - Bittorrent TRackers Automatic Checking System,” http://www.btracs.com/index_ranking.htm.
- [7] “Filesharefreak,” <http://filesharefreak.com/trackers-list/>.
- [8] “The Essential Guide To Getting Into Private Trackers,” <http://filesharefreak.com/2009/06/18/>.
- [9] “MaxMind GeoIP,” http://www.maxmind.com/app/geoip_country.
- [10] “Alexa Web Site,” <http://www.alexa.com>.
- [11] “Xunlei Website,” <http://www.xunlei.com/>.
- [12] Z. Liu, P. Dhungel, D. Wu, C. Zhang, and K. W. Ross, “Understanding and Improving Ratio Incentives in Private P2P Communities,” Submitted.
- [13] J. Falkner, M. Piatek, J. P. John, A. Krishnamurthy, and T. Anderson, “Profiling a Million User DHT,” in *Proc. IMC*, 2007.
- [14] S. A. Crosby and D. S. Wallach, “Capitalization: BitTorrent’s two Kademlia-based DHTs,” in *Technical Report TR-07-04, Rice University*, Jul. 2007.
- [15] J. A. Pouwelse, P. Garbacki, D. H. Epema, and H. Sips, “The BitTorrent P2P file-sharing system: Measurements and analysis,” in *Proc. of IPTPS*, Ithaca, NY, Feb. 2005.
- [16] A. Bellissimo, B. N. Levine, and P. Shenoy, “Exploring the use of BitTorrent as the basis for a large trace repository,” in *Tech. Rep. 04-41, UMASS Amherst*, Jun. 2004.
- [17] L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, and X. Zhang, “Measurements, analysis, and modeling of BitTorrent-like systems,” in *Internet Measurement Conference*, California, USA, Oct. 2005.
- [18] L. Guo, S. Chen, Z. X. E. Tan, X. Ding, and X. Zhang, “A performance study of BitTorrent-like peer-to-peer systems,” in *IEEE JSAC*, 2008.
- [19] G. Neglia, G. Reina, H. Zhang, D. Towsley, A. Venkataramani, and J. Danaher, “Availability in BitTorrent Systems,” in *IEEE INFOCOM*, 2007.
- [20] M. Izal, G. Urvoy-Keller, E. Biersack, P. Felber, A. Hamra, and L. Garcés-Erice., “Dissecting BitTorrent: five months in a torrent’s lifetime,” in *Passive and Active Measurements*, Apr. 2004.