

Bundling Practice in BitTorrent: What, How, and Why

Jinyoung Han
Seoul National University
Seoul, Korea
jyhan@mmlab.snu.ac.kr

Seungbae Kim
KIITC
Daejeon, Korea
sbkim@itc.kaist.ac.kr

Taejoong Chung
Seoul National University
Seoul, Korea
tjchung@mmlab.snu.ac.kr

Ted “Taekyoung” Kwon^{*}
Seoul National University
Seoul, Korea
tkkwon@snu.ac.kr

Hyun-chul Kim^{*}
Sangmyung University
Cheonan, Korea
hkim@smu.ac.kr

Yanghee Choi
Seoul National University
Seoul, Korea
yhchoi@snu.ac.kr

ABSTRACT

We conduct comprehensive measurements on the current practice of content bundling to understand the structural patterns of torrents and the participant behaviors of swarms on one of the largest BitTorrent portals: The Pirate Bay. From the datasets of the 120 K torrents and 14.8 M peers, we investigate what constitutes torrents and how users participate in swarms from the perspective of bundling, across different content categories: Movie, TV, Porn, Music, Application, Game and E-book. In particular, we focus on: (1) how prevalent content bundling is, (2) how and what files are bundled into torrents, (3) what motivates publishers to bundle files, and (4) how peers access the bundled files. We find that over 72% of BitTorrent torrents contain multiple files, which indicates that bundling is widely used for file sharing. We reveal that profit-driven BitTorrent publishers who promote their own web sites for financial gains like advertising tend to prefer to use the bundling. We also observe that most files (94%) in a bundle torrent are selected by users and the bundle torrents are more popular than the single (or non-bundle) ones on average. Overall, there are notable differences in the structural patterns of torrents and swarm characteristics (i) across different content categories and (ii) between single and bundle torrents.

Categories and Subject Descriptors

C.2.4 [Computer Communication Networks]: Distributed Systems - *Distributed Applications*

General Terms

Measurement

^{*}Corresponding Authors: Ted “Taekyoung” Kwon (tkkwon@snu.ac.kr) and Hyun-chul Kim (hkim@smu.ac.kr)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGMETRICS’12, June 11–15, 2012, London, England, UK.
Copyright 2012 ACM 978-1-4503-1097-0/12/06 ...\$10.00.

Keywords

Peer-to-Peer, BitTorrent, Content Bundling

1. INTRODUCTION

According to Sandvine’s report on Global Internet phenomena in Fall 2011, BitTorrent is responsible for a substantial amount of the Internet traffic, representing one-half to one-fifth of all the upstream traffic and one-sixth to one-fifteenth of all the downstream traffic depending on geographical regions during peak period [6]. The huge success of BitTorrent is attributed to the attractive properties of its swarming operations [16, 21, 24]. First, the swarming technique scales well even in the presence of flash crowds for popular files. Second, cooperation among peers in a swarm stimulated by the tit-for-tat incentive mechanism improve the overall system performance like throughput. Third, the tit-for-tat mechanism also addresses the free-riding problem.

This in turn has attracted the research community to investigate BitTorrent’s behavior in terms of throughput, fairness and incentive issues, revealing valuable insights into the performance aspects of BitTorrent [15, 16, 21, 24, 27]. However, most of these studies paid little attention to the internal structures of the torrents¹ (e.g., how many files are bundled in a torrent), rendering the following research questions under-appreciated by the research community: *How are torrents structured by human beings, and for what purposes? Are there any differences in the way people participate in the swarms depending on the structures of the torrents?*

We argue that understanding the structural patterns of torrents and the participant behaviors of swarms in BitTorrent with empirically-grounded evidence is important for BitTorrent stakeholders: (i) how BitTorrent service providers deal with bundling to improve the system performance, and (ii) how content providers publish torrents, especially for their financial incentives. Also, understanding and modeling how a BitTorrent-like system works from a socio-economic perspective can be linked to the research efforts in social studies and economics by addressing the following questions: “how, and for what purposes, are constituent files (component goods) bundled/packaged by people (sellers), and how peers (market consumers) respond to them?”

¹A torrent refers to a single file or multiple files that are downloaded collectively in a swarm. In this paper, a torrent and a swarm are used interchangeably; the former focuses on files, while the latter focuses on users.

To our knowledge, this measurement study is the first attempt to address the aforementioned questions with data from a large-scale BitTorrent system from the perspective of content bundling. Here, bundling [19] is a common strategy in which a publisher packages multiple files into a single torrent, which is disseminated by a single swarm, instead of disseminating individual files via separate swarms. A torrent in BitTorrent can contain either a single file or a bundle of files. If a torrent consists of a bundle of files, BitTorrent allows a user to download an arbitrary subset of the torrent.

Recently, bundling in BitTorrent has gained increasing attention, as it can mitigate the availability problem of unpopular files [18, 19] as well as reduce download times [9, 17–19, 29, 32]. However, despite the increasing interest in content bundling in BitTorrent, there have been few efforts to empirically investigate the practice of content bundling in P2P systems. This leaves researchers uncertain of how to make assumptions on bundling strategies and/or user’s accesses to bundles.

This paper seeks to bridge the gap by conducting a large scale measurement on one of the largest BitTorrent portals: The Pirate Bay [5]. We have collected datasets that contain the information on 120,550 torrents, 3,163,685 files, and 14,822,261 users, which are comprehensively analyzed. In particular, we focus on: (1) how prevalent content bundling is, (2) how and what files are bundled into torrents, (3) what motivates publishers to bundle files, and (4) how users access the bundled files.

We highlight the main contributions and key findings of this paper as follows:

1. This is the first comprehensive measurement study on content bundling practice in BitTorrent. We examine the structural patterns of and users’ behavior with bundle torrents², in comparison to single ones. We also make our developed codes and datasets online at: <http://mmlab.snu.ac.kr/traces/bundling>.
2. We find that over 72% of BitTorrent torrents contain multiple files, which indicates that bundling is prevalent.
3. We reveal that 41% of the bundle torrents consist of multiple independent yet related content files (e.g., different episodes of the same TV show), while the other 59% of the bundle torrents consist of a single main file (e.g., a movie video) and supplementary files (e.g., subtitles).
4. We find that the bundle torrents are more popular (2.1 times) than the single ones on average.
5. We observe that most files (94%) in a bundle torrent are selected by users on average.
6. To answer what motivates publishers to bundle files, we investigate top-20 publishers for each content category in single and bundle torrents, respectively, who contribute roughly 41% of all the torrents in our datasets. We observe that 68% of the bundle torrents are published by profit-driven publishers who seek financial

²Throughout this paper, a bundle torrent refers to a torrent which contains multiple files, while a single torrent contains only a single file.

gains, while only 34% of the single torrents are published by profit-driven ones. This signifies that profit-driven publishers tend to adopt the bundling strategy in BitTorrent.

7. Most of our findings lead to the following question, which provides us a socio-economic point of view on BitTorrent publishing: *what are the main incentives of content bundling in BitTorrent? How do users respond?* Our empirically grounded answers are, (i) BitTorrent publishers prefer to upload bundled torrents (72%), (ii) content bundling in BitTorrent is mainly (68%) driven by financial incentives of the publishers, which can be linked to the studies in the economics literature [7, 8, 12, 26, 30], and (iii) bundle torrents are more popular (2.1 times) than the single ones to BitTorrent users.

We organize this paper as follows. After reviewing related work in Section 2, we first present the measurement methodology in Section 3. We then show how prevalent bundling is and what files constitute torrents in Section 4. In Section 5, we analyze the characteristics of bundle torrents. We next reveal what motivates publishers to bundle files in Section 6, followed by the analysis of user access patterns in Section 7. After discussing the implications of bundling practice in Section 8, we conclude the paper in Section 9.

2. RELATED WORK

Sharing multiple torrents: Many studies on BitTorrent have focused on the advantages of users participating in multiple torrents simultaneously [11, 13, 20, 22, 31]. Note that they do not differentiate single torrents and bundle torrents. Guo *et al.* found that 85% of users concurrently access multiple torrents [13]. Yang *et al.* [31] proposed an incentive mechanism for users remaining as seeds in a subset of torrents when a user downloads multiple torrents. That is, each user calculates the aggregated download rate in a *cross-torrent* fashion in the peer selection phase, so that a user can get additional credits by participating in another torrent as a seed. Piatek *et al.* [22] also suggested a scheme that propagates peer reputations to encourage users to exchange file pieces across torrents. Sirivianos *et al.* [28] proposed a credit-based incentive scheme to stimulate more cooperation across torrents. Peterson *et al.* [20] proposed a swarm coordination system that optimally allocates the upload bandwidth of seeds among the multiple torrents to optimize the download performance. Dán *et al.* [11] explored the performance benefits achieved by dynamically merging or splitting swarms of the same content file. These studies on how to leverage multiple torrents/swarms pay no attention to whether a torrent has a single file or multiple files. In contrast, we focus on bundling, which allows peers to share multiple files in a single swarm.

Bundling in BitTorrent: Recently, Menasche *et al.* [18, 19] theoretically showed that bundling can mitigate the availability problem by combining multiple unpopular files into a single swarm. Tian *et al.* studied the performance issues in downloading files of a bundle torrent in a concurrent or sequential way with the assumption that files in a bundle torrent are highly interest-correlated [29]. Carlsson *et al.* [9] proposed a dynamic bundling strategy in which peers are assigned to download complementary contents (files or parts of files) at the time they decide to download a particular file.

Based on the dynamic bundling strategy, Zhang *et al.* [32] designed and implemented a system that can support dynamic bundling in practice. Lev-tov *et al.* [17] proposed a dynamic file selection and download strategy (among files contained in a bundle torrent) to reduce download times, where they assumed that each peer is interested in downloading only a small subset of the bundled files. While these studies were looking at potential benefits that may arise from bundling, our focus is to empirically study the current practice of bundling in BitTorrent.

Incentives of content publishing in BitTorrent: Recently, Cuevas *et al.* [10] studied the incentives of content publishing in BitTorrent from a socio-economic viewpoint, by categorizing publishers into profit-driven, altruistic, and fake ones. Here, we investigate what kind of contents are published depending on the three publisher types, and how many torrents are contributed by each publisher type from the perspective of bundling.

Bundling in economics: Product bundling is a common marketing strategy. In the economics literature, bundling strategies have been proposed as a mechanism to increase sales, extend monopoly power, and smooth demands across multiple goods [7, 12, 26, 30]. This strategy is very common in almost every business; e.g., in the cable television industry (e.g., many TV channels are often combined into a single package), in the music industry where multiple songs are combined into a single album, and in the fast food industry in which multiple foods are packaged into a combo meal. Furthermore, the bundling strategy is also widely used to promote a main product by adding supplementary items (e.g., the “free gift with purchase” concept [25] in the cosmetic industry, or a software package that contains a main software and everything it needs to operate in the software industry [23]). For the information goods that have almost zero cost to be replicated, bundling is a useful strategy to increase sales as well [8]. To distribute files as information goods, BitTorrent already supports bundling. This is the first comprehensive measurement study on the current practice of bundling in the BitTorrent ecosystem.

3. METHODOLOGY

We conduct a measurement study on one of the largest BitTorrent portal, The Pirate Bay (TPB) [5], one of the most popular torrent hosting sites. For the purpose of data collection, we developed a BitTorrent monitoring agent to keep track of each swarm by modifying the Azureus [3] client software. We also developed a torrent crawling agent to timely fetch newly released “.torrent” files by using an RSS feed³ from TPB. (We can fetch all the published torrents by using the RSS feeds from TPB.)

Figure 1 illustrates the overall measurement framework. An RSS notification of a new torrent triggers our crawling agent to immediately retrieve its publisher’s username and to request the .torrent file. By analyzing the .torrent file, the monitoring agents contact trackers through the BitTorrent Tracker protocol [2] to retrieve the lists of peers. The monitoring agents also leverage the Peer EXchange (PEX) protocol [4] to discover more peers not found via the trackers.

³TPB offers an RSS feed to announce a newly published torrent. The RSS feed provides information such as content category, content size, and publisher’s username for the new torrent.

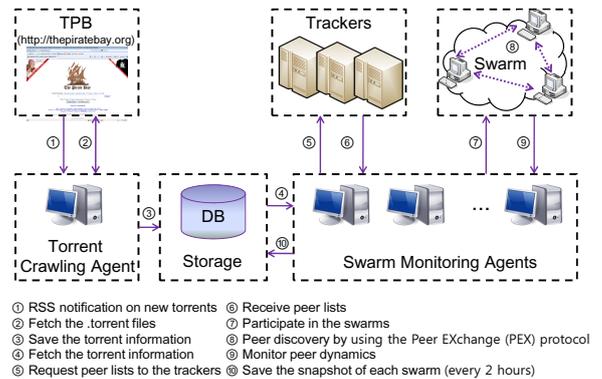


Figure 1: Measurement framework.

After finding the peers in each swarm, the monitoring agents begin monitoring each swarm. In this way, we periodically (once every two hours) obtain the snapshot of each swarm by exploiting trackers and peers. That is, the *swarm datasets* record which pieces of the file(s) are being downloaded by each peer in the swarm at the moment. For each torrent, the *torrent datasets* consist of its .torrent information, category given at TPB, publisher’s username, and published time.

3.1 Torrent Datasets

Our torrent datasets have been collected for 77 days from February 14 to May 1, 2011. The crawling agent fetched torrent data of 120,550 torrents from TPB⁴, which contains 3,163,685 files whose total volume is around 120 TB. Throughout this paper, we investigate the bundling practice of the seven major (91% and 90% in terms of the torrent counts and data volume, respectively) content categories given by TPB: Movie, TV, Porn, Music, Application, Game and E-book. The percentage of each category in terms of the number of torrents and the volume of torrents is shown in Table 1. The other torrents belonging to unknown or marginal categories are not considered in this paper.

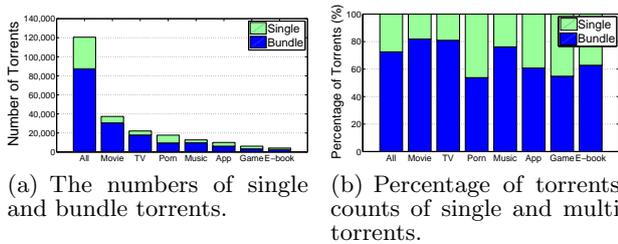
Table 1: Percentage of the number of (1st row) and the total volume of (2nd row) torrents in each category

Movie	TV	Porn	Music	App	Game	E-book
31%	18%	15%	11%	8%	5%	3%
42%	18%	13%	4%	4%	9%	1%

3.2 Swarm Datasets

For the torrents discovered between March 25 and April 26, we have periodically (once every two hours) captured swarm snapshots, to investigate access pattern of peers participating in the swarms. We restrict the swarm dataset collection and analysis to those of the torrents collected between March 25 and April 26, due to the performance limitations of our monitoring facilities, which consist of 14 desktop PCs (admittedly research-grade). Consequently, we have captured swarm snapshots of 43,837 torrents, in which 14,822,261 peers were exchanging 1,301,354 files.

⁴The average number of daily published torrents in TPB is around 1.6 K.



(a) The numbers of single and bundle torrents.

(b) Percentage of torrents counts of single and multi torrents.

Figure 2: Bundling is widely used in BitTorrent.

Notice that the torrent dataset is used in Sections 4, 5, and 6 for investigating bundling patterns from the perspective of publishers, while the swarm dataset is used in Section 7 for investigating the bundling patterns and user access patterns together from the perspectives of publishers and users.

4. SINGLE VS. BUNDLE

In this section, we seek to answer the following questions: (1) to what extent content bundling is prevalent in BitTorrent, and (2) how and what files are bundled into torrents, particularly in terms of the number, volume, and types of files. Note that we focus on the comparison between single and bundle torrents in this analysis. Recall that the torrent datasets of 120,550 torrents are used here.

4.1 Bundling is widespread

To analyze how prevalent content bundling is in BitTorrent, we compare bundle torrents with single ones in terms of the number of torrents and the volume of torrents. Note that the volume of torrents is the total size of files of the given torrents. Figure 2 shows that over 72% of the torrents contain multiple files, which means content bundling is widely used. In the Music category, around 80% of the torrents use bundling, which indicates that BitTorrent users often share a collection of music files from the same genre, player, composer, or album. Likewise, over 80% of the torrents in the Movie category use bundling, mostly because users often package: (i) multiple movie files of the same series (e.g., sequels), or (ii) a main video file and other supplementary files like subtitles. Meanwhile, around 60% of the torrents in the Application and Game categories contain multiple files (e.g., installation files and subsidiary files such as how-to documents), while the other torrents (40%) have a single installation file.

We next compare the volume of single and bundle torrents in Figure 3. The average size of a bundle torrent (1.2 GB on average) is approximately twice as large as that of a single one (0.6 GB on average) as shown in Figure 3(a). In the Movie category, the average size of a bundle torrent (1.4 GB) is not so higher than that of a single one (1.2 GB), while the average size of a bundle torrent in the other categories is significantly higher (mostly more than 2 times) than that of a single one. This is because a Movie bundle torrent usually consists of a large video file with smaller supplementary files like subtitles, to be detailed in Section 5. Figures 3(b) and 3(c) show that the volume of bundle torrents outweighs that of single ones; the volume of bundle torrents across the seven categories accounts for over 80% of the volume of total torrents that we investigated. The volume of all the torrents reaches around 120 TB.

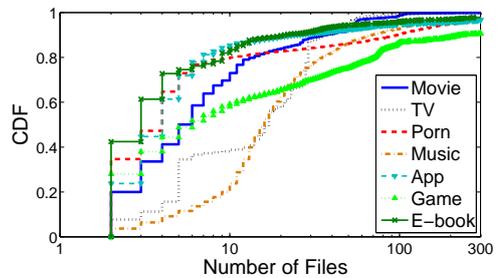


Figure 4: Number of files in a bundle torrent.

4.2 How files are bundled

In this subsection, we investigate the number of files and their file extensions (e.g., avi, jpg, mp3) in single and bundle torrents to understand the internal structures of the torrents. Our torrent dataset consists of 120,550 torrents, which contain 3,163,685 files with 5,538 different file extensions. Among them, the total number of files contained in 87,430 bundle torrents is 3,130,565; i.e., 36 files per bundle torrent on average.

The number of files in a bundle torrent: Figure 4 shows the CDF of the number of files in a bundle torrent across different categories of torrents. As shown in Figure 4, bundle torrents of the Music category contain significantly more files than those of the other categories; around 80% of Music bundle torrents contain more than 10 files. In contrast, 73% of E-book bundle torrents contain 4 files or less. Around 70% of TV bundle torrents contain 5 to 30 files because users often package and share the series of the same TV drama (e.g., 22 episodes of “*Gossip Girl*” season-3 or 24 episodes of “*24*” season-4). Note that in the Porn bundle torrents, around 80% contain 7 files or less, while over 10% contain 100 files or more.

File extension analysis: Table 2 shows the top 3 file extensions in the single and bundle torrents, respectively. In the Movie category, although the number of .avi files in bundle torrents is placed third, the volume of the .avi files ranks top. Note that the file extension .txt ranking second in the Movie bundle torrents is often used for promoting web sites to be discussed in Section 6. In the Porn category, around 92% of files contained in bundle torrents are with .jpg extension, which means publishers of those files often bundle a lot of pornographic pictures into a torrent. However, video files account for more than 90% of the total volume of all files contained in the Porn bundle torrents, since the size of video files are much larger than that of images. Notice that most of the top 3 file extensions in terms of the number and the volume of files in the single torrents in the Movie, TV, and Porn categories are those of video files. The bias toward the video files in these three categories is also significant in the bundle torrents when we look at the volume percentages of file extensions in Table 2. In the Music and E-book categories, .mp3 files and .pdf files are dominant in terms of both number and volume, respectively. Interestingly, the .avi file extension ranks second in terms of volume in Music single torrents, which are music-video files. In the Application and Game categories, there are numerous file extensions; however, the volume of .exe, .iso, and .rar files is dominant. Note that most of the high rank file types are

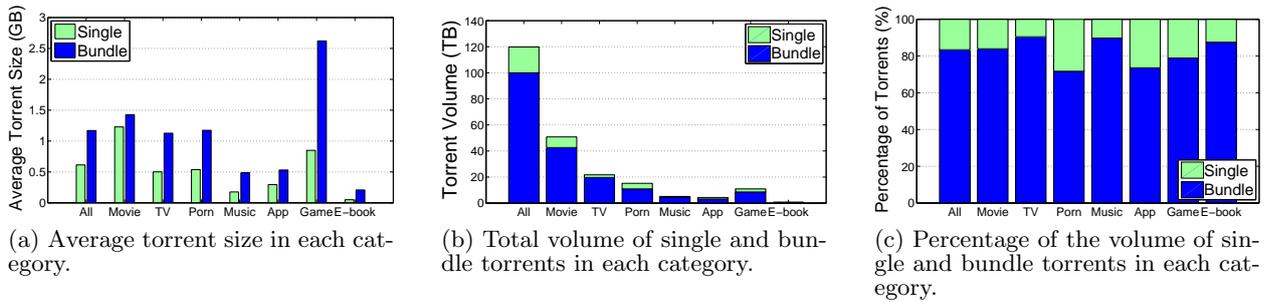
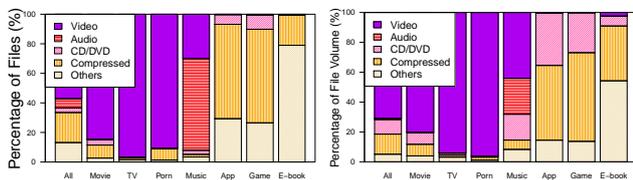


Figure 3: Volume of bundle torrents is substantially larger than that of the single ones.

Table 2: Top 3 file extensions in single (S) and bundle (B) torrents in terms of number and volume, respectively

Type		S-Rank1(%)	S-Rank2(%)	S-Rank3(%)	B-Rank1(%)	B-Rank2(%)	B-Rank3(%)
All	Number	avi(32)	rar(14)	wmv(14)	jpg(24)	mp3(12)	rar(10)
	Volume	avi(37)	wmv(15)	rar(11)	avi(36)	rar(14)	vob(9)
Movie	Number	avi(68)	rar(9)	mkv(8)	rar(30)	txt(11)	avi(9)
	Volume	avi(61)	wmv(14)	iso(8)	avi(41)	vob(15)	rar(14)
TV	Number	avi(65)	mkv(16)	mp4(13)	rar(47)	avi(11)	txt(8)
	Volume	avi(52)	mkv(26)	mp4(13)	avi(46)	rar(19)	mkv(16)
Porn	Number	wmv(49)	avi(21)	mp4(15)	jpg(92)	avi(2)	wmv(2)
	Volume	wmv(53)	mp4(19)	avi(19)	avi(49)	wmv(31)	mp4(7)
Music	Number	mp3(59)	avi(15)	mp4(9)	mp3(77)	jpg(6)	wma(3)
	Volume	mp3(24)	avi(17)	iso(15)	mp3(47)	vob(27)	avi(5)
App	Number	rar(45)	exe(22)	zip(17)	wav(5)	jpg(5)	exe(4)
	Volume	rar(39)	iso(35)	zip(8)	iso(39)	rar(12)	exe(10)
Game	Number	rar(46)	exe(21)	zip(13)	wav(11)	png(5)	rar(3)
	Volume	rar(48)	iso(27)	exe(10)	rar(24)	iso(16)	mpq(5)
E-book	Number	pdf(70)	rar(14)	zip(5)	pdf(28)	jpg(18)	doc(7)
	Volume	pdf(50)	rar(22)	zip(14)	pdf(63)	rar(11)	djvu(5)

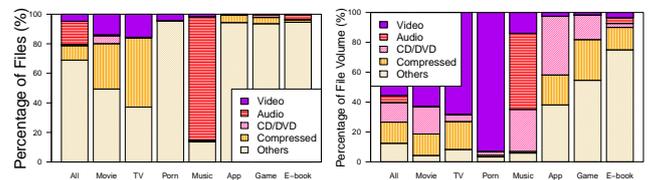


(a) Percentage of the number of file types in single torrents. (b) Percentage of the volume of file types in single torrents.

Figure 5: Percentage of file types for single torrents in terms of number and volume.

linked to the main files of the torrents to be discussed in Section 5.

Video, audio, cd/dvd, and compressed files analysis: From the above results, we note that video, audio, cd/dvd, and compressed files are major file types constituting torrents. Thus, we plot the percentage of the numbers and volume of the file types in single and bundle torrents in Figures 5 and 6, respectively. Notice that we consider compressed files as single files and do not investigate what files are contained in the compressed files because it is not easy to identify what files are contained in the compressed files until we actually download them, which might be illegal. As shown in Figures 5 and 6, the numbers of compressed files



(a) Percentage of the number of file types in bundle torrents. (b) Percentage of the volume of file types in bundle torrents.

Figure 6: Percentage of file types for bundle torrents in terms of number and volume.

account for approximately 20% and 10% across all the categories of the single and bundle torrents, respectively. In the Application and Game categories, notice that over 60% of the single torrents in terms of number are compressed mainly for installation purposes. Along the same line, the volume of cd/dvd images is substantial in both single and bundle torrents. Figures 5(b) and 6(b) show that the volume of video files accounts for over 70% in the single torrents but less than 60% in the bundle ones. The number and volume of video files are dominant in the Movie, TV, and Porn categories in the single torrents. However, the number and volume of video files in Movie and TV categories in the bundle torrents

show a less portion than those in the single torrents. Note that the number of video files in Porn bundle torrents shows a small portion compared to that of video files in single ones because porn publishers often bundle a lot of pornographic pictures into a torrent. Interestingly, (music) video files account for around 45% of the total volume of all files in the Music single torrents.

5. MAIN FILE ANALYSIS IN BUNDLING

In this section, we investigate how many “main files” are included in each bundle torrent. Here a main file is the primary media file in a torrent; other supplementary files in the torrent are not counted as main files. For example, an .avi file in a Movie torrent or an .mp3 file in a Music torrent is the main file. Note that there may be multiple main files in a torrent; if a publisher bundles two episodes of a TV drama, there are two main files in the torrent.

5.1 Identifying Main Files

When we identify main files, we first consider the torrent categories and file extensions as shown in Table 3. In addition, we examine file names to further refine our identification; for instance, a file whose name contains special keywords such as “sample” or “trailer” is excluded.

We then count the number of main files in each torrent. A torrent containing k main files is denoted by “bundle- k .” Bundle- k ($k > 1$) can be linked to a bundling strategy that packages multiple products into a single combined product. On the other hand, bundle-1 can be linked to another bundling strategy that promotes a main product by adding supplementary items (e.g., the “free gift with purchase” concept [25] in the cosmetic industry). As mentioned earlier, we exclude torrents with the compressed files. We also exclude the torrents in the Game and Application categories because it is difficult to identify their main files. Note that a bundle-1 torrent contains a main file with other supplementary files, while a single torrent has only a single file.

Table 3: File extensions identified as main files in five categories

Category	File extensions
Movie	avi, wmv, mkv, mp4, vob, mpg, mov, ...
Porn	jpg, avi, asf, wmv, mpg, mkv, gif, ...
TV	avi, mkv, mp4, mpg, mov, m4v, wmv, ...
Music	mp3, m4a, wma, flac, m3u, mp4, wav, ...
E-book	pdf, epub, mobi, dvju, ps, ...

5.2 Constituents of Bundle- k

Bundle-1 versus bundle- k : Figure 7 shows the percentage of bundle-1 and bundle- k ($k > 1$) torrents in all the bundle torrents in terms of the number and volume of torrents. As shown in Figure 7(a), the bundle-1 torrents account for 59% across all the categories, which means users often use a torrent to share only a single main file. However, in the Music category, around 92% of the torrents have multiple music files, which means a Music bundle torrent is mostly used to offer several music files as one combined product like a single music album. The number of bundle-1 torrents in the Movie category accounts for around 88% of that of all the Movie bundle torrents, which indicates that users often

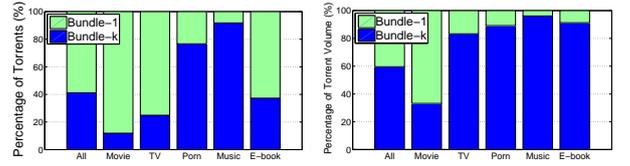


Figure 7: Comparison between bundle-1 and bundle- k ($k > 1$) torrents.

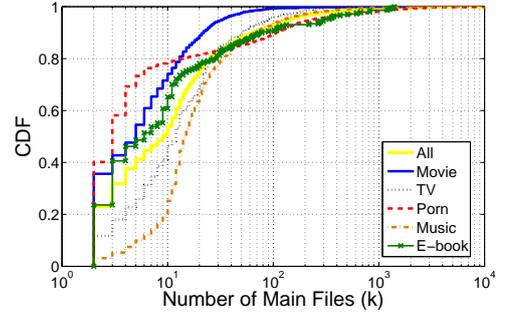


Figure 8: CDF of bundle- k torrents ($k > 1$).

share a single movie file with supplementary files such as subtitles. Figure 7(b) shows that the volume of bundle- k ($k > 1$) is around 60% on average. Note that the volume of bundle- k ($k > 1$) torrents in the TV category accounts for around 83% of that of all the TV bundle torrents even though the number of bundle- k ($k > 1$) torrents is less than 25%. This is because users often bundle a large number of episodes into a bundle- k ($k > 1$) torrent in the TV category.

The number of main files in a bundle- k ($k > 1$): To analyze how many main files are included in a bundle torrent depending on the categories, we plot the CDF of bundle- k ($k > 1$) in Figure 8. Porn torrents exhibit interesting phenomena; 40% of torrents have two main files (mostly video files), and 11% of the torrents have 100 main files or more (mostly image files), which explains the heavy tail distribution. E-book torrents have a similar but weak pattern to the Porn ones; 24% of torrents contain only two main files and 9% of torrents have 100 main files or more. In the Movie category, around 74% of torrents have 10 main files or less, which reflects that a movie series normally has less than 10 episodes. On the other hand, 60% of the TV torrents have 10 episodes or more since a TV series typically consists of more than 10 episodes. Similarly, the number of main files in a Music torrent falls between 10 and 100 with 74% probability.

A torrent size versus the number of main files in a bundle- k : To analyze the correlation between the torrent size and the number of main files (i.e., k of bundle- k) in a torrent, we adopt Pearson’s correlation coefficient [14], denoted by ρ . Figure 9 shows the torrent sizes of bundle- k torrents as k increases. The ρ values of the Movie, TV, Porn, Music, and E-book categories are 0.372, 0.746, 0.056, 0.434, and 0.501, respectively. Except for Porn category, a positive correlation between the torrent size and the number of main

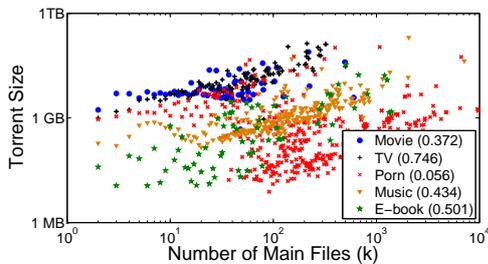


Figure 9: The torrent size and k is highly correlated in all the categories except for Porn.

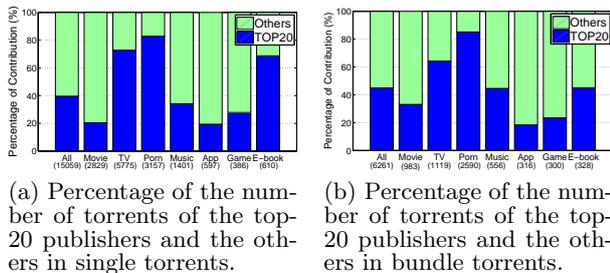


Figure 10: Top-20 publishers contribute a significant number of torrents depending on each content category. The numbers below each category indicates the number of torrents published by the top-20 publishers.

files is found in all the categories. Especially, the correlation is relatively stronger in the TV, Music, and E-book categories because files are typically of the similar size; e.g., an avi file of a TV drama is around 700 MB, an mp3 file of a song is around 10 MB, and a pdf file of an E-book is around 2 MB. The correlation in the Movie category is somewhat weaker since the sizes of .vob files vary diversely according to their included functions like DVD menu or captions. Note that there is a negligible correlation in the Porn category because pictures and videos are two major disparate elements, which are mixed into a torrent without any regular pattern.

6. PUBLISHER ANALYSIS

In this section, we study publishers in BitTorrent from a socio-economic point of view by unravelling who publishes bundle or single torrents and why. To this end, we first divide publishers into three types based on the purposes of content publishing [10]: (1) *fake* publishers who publish fake contents, (2) *profit-driven* publishers who usually promote their own web sites for financial gains like advertising, and (3) *altruistic* publishers. We also investigate the level of contribution (i.e., the number of published torrents) of publishers in each type and whether and how they publish torrents across multiple content categories.

6.1 Classifying Publishers

Fake publishers publish “fake” contents. They often inject malware into files or make garbage files with catchy titles (e.g., recently released popular movies such as “The Green Hornet” or “Black Swan” as of April, 2011) or intriguing titles (e.g., adult movies); these publishers appear to

be malicious users to disseminate malware or be associated with antipiracy agencies to hinder the distribution of copyrighted content [10]. The administrators of TPB remove the accounts of publishers and their published torrents on TPB when they are reported as fake publishers from other users.

Profit-driven publishers publish contents for financial incentives. They often promote one or more web sites with financial incentives [10]. Profit-driven publishers usually use major BitTorrent portals such as TPB as a platform to advertise their profitable web sites (e.g., BitTorrent portals that are associated with private trackers or adult sites) to users. For this purpose, they publish popular torrents where they attach URLs of their web sites in various manners: (i) textbox in the web page associated with each published content, (ii) title of a text file (mostly .txt, .nfo, and .html files), (iii) title of a .torrent file, and (iv) title of a main file.

Altruistic publishers publish contents only for sharing. They neither promote any web site nor distribute fake contents.

In order to systematically classify publishers, we take following steps:

1. **Checking publisher’s account:** We first check whether the account of each publisher is removed from TPB or not after the observation period. If the account is removed and their associated torrents uploaded by the publisher are not available on TPB, we conclude the publisher is a fake one. We double check the account of the publisher in the Suprbay forum⁵ where users report fake publishers on TPB.
2. **Checking web pages:** When the account is available, we next examine the textbox in the web page associated with each published torrents. If there are any URLs on the textbox, we identify them as profit-driven publishers.
3. **Checking files:** We then examine whether any file in a torrent has a URL information to be advertised. If so, we classify them as profit-driven publishers. We also investigate the titles of torrents. If there are any URLs to advertise specific web sites or private trackers embedded in the titles, the publishers are classified as profit-driven ones.
4. The remaining publishers are classified as altruistic publishers because they do not seem to promote any URL nor upload fake contents.

6.2 Contribution of Top-20 Publishers

Using the above methodology, we classify the top-20 publishers in terms of the number of published torrents for each content category in single and bundle torrents, respectively, into three publisher types. There would be total $20 \times 7 \times 2 (= 280)$ publishers; however there are total 242 publishers since some publishers are overlapping. Although the top-20 publishers in the single torrents account only for 2.7%, 7.0%, 9.6%, 4.6%, 1.8%, 2.6%, and 6.7% of the total number of publishers in the Movie, TV, Porn, Music, Application, Game, and E-book categories, respectively, they contribute a substantial portion of torrents as shown in Figure 10(a). Likewise, the top-20 publishers in the bundle torrents account only for 1.0%, 3.2%, 9.7%, 1.7%, 1.4%, 2.6%, and

⁵<https://forum.suprbay.org>

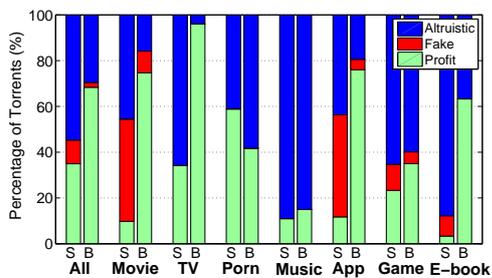


Figure 11: Percentage of the number of torrent of each publisher type in single (S) and bundle (B) torrents.

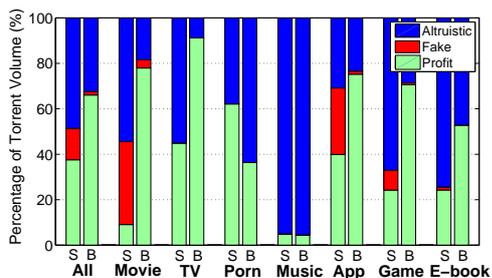


Figure 12: Percentage of the volume of torrent of each publisher type in single (S) and bundle (B) torrents.

10.5% of the total number of publishers in the Movie, TV, Porn, Music, Application, Game, and E-book categories, respectively, but they contribute a significant portion of torrents as well as shown in Figure 10(b). Note that the top-20 publishers in the bundle torrents contribute around 73%, 83%, and 69% of all the bundle torrents in the TV, Porn and E-book categories, respectively. Overall, the top-20 publishers across the seven categories contribute roughly 41% of all the single and bundle torrents in our torrent datasets.

Figures 11 and 12 show the percentage of contribution for each publisher type in terms of number and volume of all the torrents, respectively. Interestingly, the contribution of the profit-driven publishers is significant (around 68%) in bundle torrents while the contribution of the profit-driven publishers in single torrents accounts only for around 34%; profit-driven publishers tend to prefer to upload the bundle torrents. Especially, the percentage of the number of torrents of the profit-driven publishers is higher in the Movie (75%), TV (96%), Application (76%), and E-book (63%) categories. This is because profit-driven publishers often use the additional text files such as .txt, .nfo, and .html files to promote their web sites. When we look at the bundle torrents of profit-driven publishers only, the numbers of .txt files account for 11%, 9%, 11%, and 16% of the files in the bundle torrents in these four categories, respectively. Note that .txt file extension of the bundle torrents ranks third (5%) in terms of the number of file extensions across seven categories. However, the torrents in the Porn category exhibit a different pattern; the percentage of contribution of profit-driven publishers in single torrents is higher than the one in bundle torrents since they usually use the textbox in the web pages to advertise their URLs instead of using additional text files.

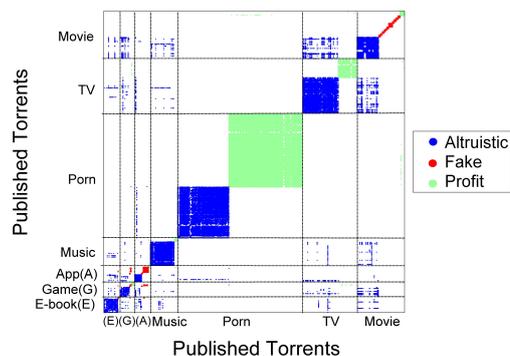


Figure 13: Cross-category publishing of each publisher in single torrents.

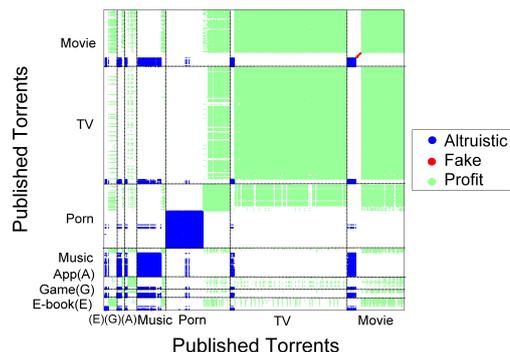


Figure 14: Cross-category publishing of each publisher in bundle torrents.

Fake publishers exhibit also interesting patterns. As shown in Figures 11 and 12, the percentage of their contribution in single torrents is higher than the one in the bundle torrents; they tend to prefer to upload single torrents. This is because fake publishers have no reason to bundle additional text files; they just inject “fake” files instead of original files. Note that the contribution of the fake publishers (especially in the single torrents) in the Movie and Application categories is higher than the ones in the others since they try to hinder the distribution of copyrighted movie contents and to disseminate malware, respectively. We also find that file extensions published by fake publishers are mostly .avi, .rar, and .exe files. This is because fake publishers try to attract users by .avi movie files with catchy titles in Movie torrents or infect users by .exe executable files in Application torrents. Also .rar files are widely used to hide their fake contents both in the Movie and Application categories.

In summary, we show that a significant amount (68%) of bundling is done by profit-driven publishers. From this result, we conclude that bundling in BitTorrent is mainly driven by financial considerations, which can be linked to the bundling in economics [7, 12, 26, 30]. In other words, the publishers who have financial considerations often adopt the bundling strategy in BitTorrent. Later, we will investigate how users respond to the bundling practices in Section 7.

6.3 Cross-category Publishing of Top-20 Publishers

In this subsection, we examine whether and how the top-

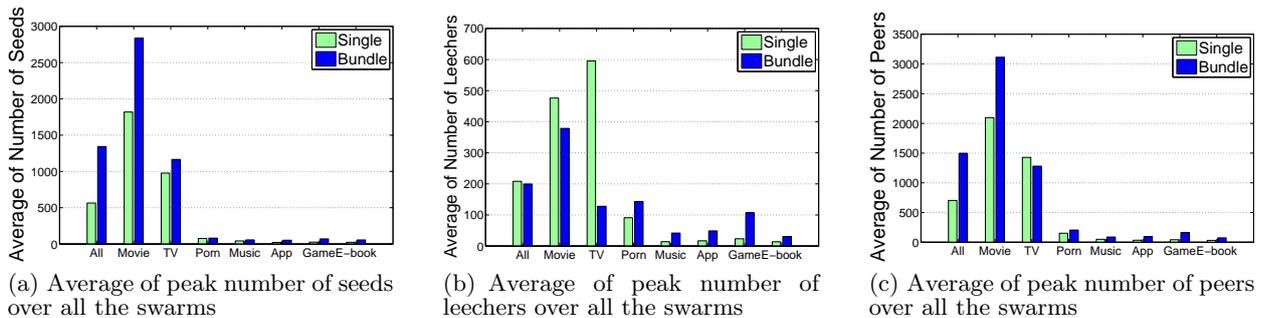


Figure 15: Bundle swarms are more popular than single ones on average.

20 publishers publish torrents across multiple content categories in single and bundle torrents, respectively. To this end, we examine the number of published torrents across the different content categories in our torrent datasets.

Figures 13 and 14 show whether a publisher, who published a torrent in a content category, also publishes another in the other categories. That is, a color pixel at (x, y) represents whether the publisher of torrent x in the horizontal axis also publish torrent y in the vertical axis. We enumerate each torrent according to the seven categories and three publisher types in the horizontal and vertical axes. Figure 14 shows that profit-driven publishers in bundle torrents often publish torrents in other content categories. For example, a profit-driven publisher in the TV category has a strong tendency of publishing torrents in the Porn and Movie categories as well. On the contrary, profit-driven publishers in single torrents mostly publish torrents only in a single category (mostly Porn) as shown in Figure 13 because profit-driven publishers who publish single Porn torrents often use the textbox in the webpage rather than the advertisement text files. Interestingly, altruistic publishers of the single and bundle torrents in the Porn category mostly focus on the Porn category alike while the other altruistic publishers (not in the Porn category) publish torrents across multiple categories. This indicates that altruistic Porn publishers are solely interested in uploading the Porn contents.

7. USER ACCESS PATTERN ANALYSIS

In this section, we address the following questions. Do users prefer to download bundle torrents over single ones? Are bundle torrents more available than single ones in practice? Is there a correlation between the popularity and the number of main files? Do users actually prefer to download all the files in a bundle torrent? To answer these questions, we investigate user access patterns in terms of multiple metrics: (1) popularity is the number of seeds, leechers, and peers (i.e., both seeds and leechers) in a swarm at peak time, (2) area popularity is the summation of the periodically sampled numbers of peers during the swarm’s lifetime, (3) availability is the average of the sum of the number of seeds and the fraction of the file(s) in leechers in a swarm during its lifetime [1], (4) seed ratio is the fraction of time with at least one seed available over the swarm’s lifetime [19], and (5) file selection ratio is the ratio of the number of files requested by users to the number of all the files in a bundle torrent. Here, a swarm’s lifetime is from the moment of the first seed to the moment of the last seed and no seed appears

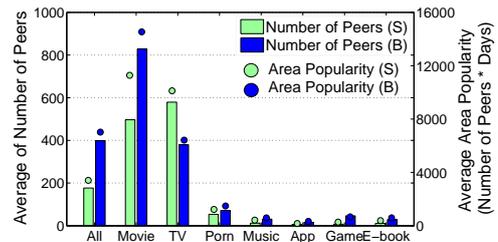


Figure 16: Average of number of peers during swarm’s lifetime and average area popularity of single (S) and bundle (B) swarms in each category.

after the moment. As we focus on user behaviors here, single swarms and bundle swarms are mostly used instead of single torrents and bundle torrents, respectively. Note that we use the 33 days’ swarm datasets of the 43,837 torrents collected from March 25 to April 26 for this analysis, as explained in Section 3.2.

7.1 Popularity Analysis

We first measure the popularity of the single and bundle swarms in terms of the number of seeds, leechers, and peers (i.e., both seeds and leechers) at peak time in Figure 15. Figure 15(a) shows the average of the peak number of seeds of all the swarms in each category. Notice that the average of the seed popularity of bundle swarms is around 2.4 times as large as that of the single ones. In other words, the bundle swarms are likely to be more available, and hence support faster download. Note that the average of the seed popularity in Movie bundle swarms reaches almost 3000. Figure 15(b) shows the average of the peak number of leechers of all the swarms in each category. Notice that the averages of the leecher popularity of single swarms in the Movie and TV categories are higher than those of bundle swarms. Interestingly, the average of the leecher popularity of the TV single swarms is 4.7 times as large as that of the TV bundle ones. This is due to the timing-sensitive nature of the TV drama; a single torrent containing a new episode aired this week is usually much more popular than a bundle torrent containing multiple old episodes of the same TV drama. On the other hand, the average of the leecher popularity of single swarms in the other categories (except for TV and Movie) is smaller than that of bundle ones. Finally, Figure 15(c) shows that the average of the peer popularity of bundle swarms is 2.1 times larger than that of the single ones.

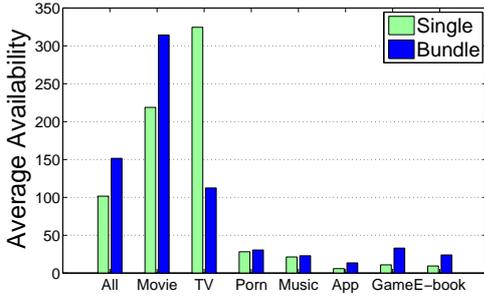


Figure 17: Average of availability of single and bundle swarms in each category.

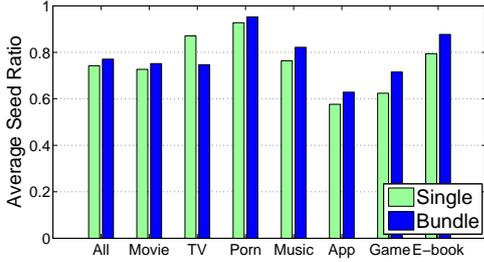


Figure 18: Average of seed ratio of single and bundle swarms in each category.

Additionally, we estimate the “Area Popularity” which reflects not only the instantaneous popularity but also the popularity over time. The *Area Popularity* of swarm s can be calculated as follows:

$$\text{Area Popularity}(s) = \sum_{\text{lifetime of } s} P(s, t)$$

where $P(s, t)$ is the periodically measured number of peers in swarm s at time t . Figure 16 shows that the average of number of peers during swarm’s lifetime and area popularity in bundle swarms are larger than those in single ones across all the categories except for TV, which is consistent with Figure 15(c).

7.2 Availability Analysis

We then analyze the availability of single and bundle swarms. While some studies (e.g., [19]) define the availability as the existence of seeds, our definition also accounts for the union of file pieces that leechers have in a swarm. For example, if there are two seeds and the portion of pieces of leechers in a swarm is 75% of the whole content file(s), the availability is 2.75. Many BitTorrent clients also use this definition [1]. Figure 17 shows that the bundle swarms are mostly more available than the single ones except for TV, which is in line with the average of number of peers during swarm’s lifetime in Figure 16.

Finally, we show the seed ratio, the fraction of time with at least one seed available over the swarm’s lifetime [19] in Figure 18. The seed ratio of the bundle swarms is slightly higher than that of the single ones across all the categories except for TV, which is somewhat in line with Figure 17. Interestingly, the seed ratio of the Porn category is significantly high (around 0.95).

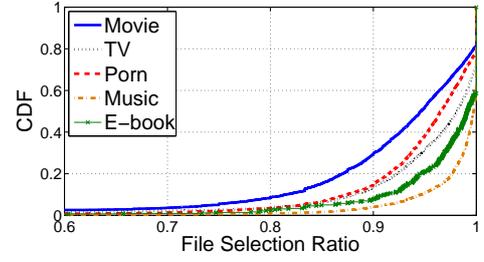


Figure 19: Ratio of the number of files requested in a bundle swarm. Over 94% of all the files in a bundle torrent are selected on average.

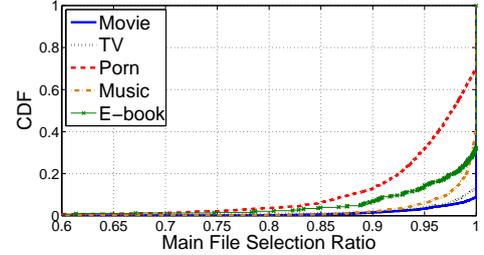


Figure 20: Ratio of the number of main files requested in a bundle swarm. Over 98% of all the main files in a bundle torrent are selected on average.

7.3 The Number of Files Requested by Users in a Bundle Torrent

We next analyze the ratio of the number of files requested by users in a bundle torrent. In the BitTorrent software, users can select any subset of the bundled files. Figures 19 and 20 show the ratio of the numbers of files and main files requested by users to the numbers of all the files and all the main files in a bundle torrent, respectively. As shown in Figure 19, over 94% files in a bundle torrent are selected on average by users. Furthermore, almost every main file (over 98% on average) in a bundle torrent is selected by users in Figure 20. Especially, all the files are selected in 42% of bundle torrents in the Music category. On the other hand, all the files are selected in only 19% of bundle torrents in the Movie category. This is because users often deselect non-main files in Movie bundle torrents. Interestingly, however, all the main files in 91% of Movie bundle torrents are selected by users; users prefer to download all the main video files because they are usually related (e.g., Shrek 1 and Shrek 2). The ratio of the main files requested by users in the Porn category is relatively lower than that in the other categories as shown in Figure 20; (i) main files in the same torrent are often disparate like video and image and (ii) the relation of main files is usually weak. Overall, the selection ratio of the main files is higher than that of all files (i.e., main and non-main files) across all the categories.

7.4 Swarm Behaviors versus Bundle- k

Popularity versus number of main files: We first analyze the correlation between the number of main files (i.e., k in bundle- k) in a torrent and the number of peers in its swarm at peak time. Figure 21 shows the number of peers that access a bundle- k torrent at peak time as k in-

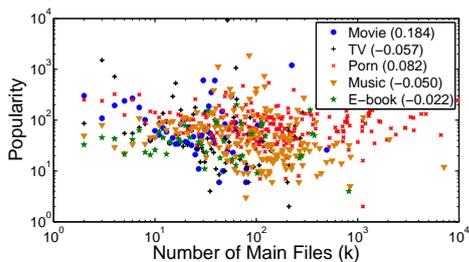


Figure 21: Popularity versus Bundle- k .

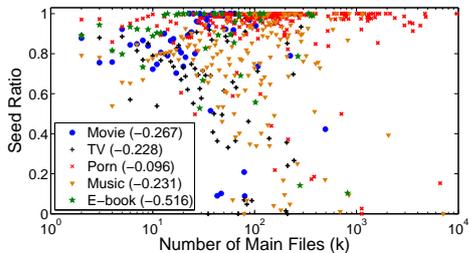


Figure 22: Seed ratio versus Bundle- k .

creases. To quantify the correlation, we again calculate the Pearson’s correlation coefficient ρ . As shown in Figure 21, the ρ values of the Movie, TV, Porn, Music, and E-book categories are 0.184, -0.057, 0.082, -0.050, and -0.022, respectively. The correlation in the Movie category is slightly positive since users tend to download a torrent containing multiple episodes of the same movie series. However, k in bundle- k torrents has no strong correlation with the popularity of the torrent across all the categories except for Movie.

Seed ratio versus number of main files: We then analyze the correlation between k and the seed ratio of the swarm. Figure 22 shows the seed ratio as k increases. The ρ values of the Movie, TV, Porn, Music, and E-book categories are -0.267, -0.228, -0.096, -0.231, and -0.516, respectively. Figure 22 is not in line with Figure 21. Overall, there is clear negative correlation between the seed ratio and k except for Porn. This is because seeds may not prefer to stay in the swarm since the seeding overhead worsens as the torrent size increases, which is consistent with Section 5 (See Figure 9). However, there is negligible correlation in the Porn category as well since correlation between the torrent size and k is insignificant as shown in Figure 9.

Main file selection ratio versus number of main files: We finally analyze the correlation between k and the ratio of main files requested by users in a bundle- k torrent. As shown in Figure 23, the ρ values of the Movie, TV, Porn, Music, and E-book categories are -0.002, -0.240, -0.056, -0.398, and -0.478, respectively. Interestingly, there is significant negative correlation in the TV, Music and E-book categories; users prefer not to download all the main files in a bundle- k as k increases, but to download only a small subset of interested main files such as recent TV episodes and personally favorite songs. On the contrary, there is no clear correlation in the Movie category. This is mainly because all the main files are selected by users in 91% of Movie bundle torrents as shown in Figure 20. Note that there is also negligible correlation between the features of Porn torrents and k when we look at Figures 9, 21, 22, and 23. This implies

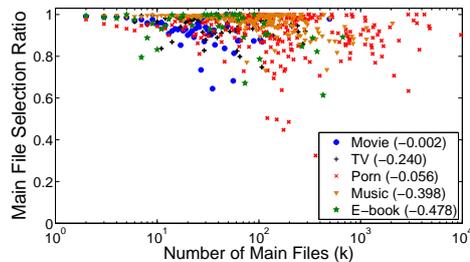


Figure 23: Main file selection ratio versus Bundle- k .

that porn users tend to be indifferent to the number of main files.

8. DISCUSSIONS

This paper is motivated by the following question: *How can we understand bundling in BitTorrent?* Specifically, we investigate the bundling practice in BitTorrent from the following perspective: Are similar patterns to the bundling in economics observed in bundling in BitTorrent? Our empirically grounded answer is that we can observe a similar trend between the bundling in economics [7, 8, 12, 26, 30] and the bundling in BitTorrent: (i) bundling in BitTorrent is mainly (68%) driven by financial considerations, and (ii) current bundling practice in BitTorrent adopts both of the bundling cases: (a) multiple comparable products are combined into a single package, and (b) supplementary items are added to a main product. We believe that the analysis of the bundling practice can give insights to content providers for making their marketing/publishing strategies. For instance, using bundling with additional text files to advertise their profitable web sites can be an advertising business model to the content providers, which is similar to the marketing strategy using leaflets.

Some theoretical studies [9, 17–19, 29] suggest that bundling can improve the content availability and reduce download times. However, to exploit the above advantages in real environment, P2P/BitTorrent service providers need to consider how they effectively constitute bundles and to predict how users participate in the swarms of the bundle. We expect that our empirical analysis of the bundling practice can be a reference for P2P/BitTorrent service providers. For example, a small number of files (say, less than 15) need to be bundled into a torrent in the E-book category if the average seed ratio is required to be over 85% since the seed ratio significantly decreases as the number of main files increases (See Figure 22). Another example is that if a P2P/BitTorrent service provider wants to maintain the main file selection ratio over 98% in the TV category, less than 10 files need to be bundled into a torrent (See Figure 23).

While the bundling strategy is mainly adopted by firms in traditional businesses for increasing sales, e.g., extending monopoly power and smoothing demands across multiple goods [7, 12, 26, 30], this paper shows that bundling is actively adopted by anonymous publishers as well. Interestingly, there are notable similarities between the bundling in BitTorrent and traditional businesses. Further studies on the links between the BitTorrent and business practices may be interesting.

9. CONCLUSIONS

We conducted comprehensive measurements on the bundling practice to understand the structural patterns of torrents and the participant behaviors of swarms in BitTorrent. From the datasets of the BitTorrent files and swarm dynamics, we analyzed: (1) how prevalent content bundling is, (2) how and what files are bundled into torrents, (3) what motivates publishers to bundle files, and (4) how peers access the bundled files. We first found that bundling is widespread for file sharing. We observed that 41% of the bundle torrents consist of multiple main files, while the other 59% of the bundle torrents consist of a single main file and supplementary files. We also observed that most files (94%) in a bundle torrent are selected by users and the bundle torrents are more popular than the single ones on average. We further revealed that bundling in BitTorrent is mainly (68%) driven by financial incentives, which can be linked to the bundling in economics.

10. ACKNOWLEDGMENTS

We thank our shepherd Anirban Mahanti and anonymous reviewers for their help and invaluable comments. This work was supported by the KCC (Korea Communications Commission), Korea, under the R&D program supervised by the KCA (Korea Communications Agency) (KCA-2012-11-911-05-002). The ICT at Seoul National University provided research facilities for this study.

11. REFERENCES

- [1] Availability, vuze. <http://wiki.vuze.com/w/Availability>.
- [2] Bittorrent tracker protocol. http://wiki.theory.org/BitTorrent_Tracker_Protocol.
- [3] Open sourced bittorrent client, vuze. <http://www.vuze.com>.
- [4] Peer exchange. http://en.wikipedia.org/wiki/Peer_exchange.
- [5] The pirate bay. <http://thepiratebay.org/>.
- [6] Sandvine global internet phenomena report: Fall 2011. http://www.sandvine.com/news/global_broadband_trends.asp.
- [7] W. J. Adams and J. L. Yellen. Commodity bundling and the burden of monopoly. *Quarterly Journal of Economics*, 90(3):475–498, 1976.
- [8] Y. Bakos and E. Brynjolfsson. Bundling information goods: Pricing, profits, and efficiency. *Management Science*, 45(12):1613–1630, 1999.
- [9] N. Carlsson, D. L. Eager, and A. Mahanti. Using torrent inflation to efficiently serve the long tail in peer-assisted content delivery systems. In *IFIP Networking*, 2010.
- [10] R. Cuevas, M. Kryczka, A. Cuevas, S. Kaune, C. Guerrero, and R. Rejaie. Is content publishing in bittorrent altruistic or profit-driven? In *ACM CoNEXT*, 2010.
- [11] G. Dán and N. Carlsson. Dynamic swarm management for improved bittorrent performance. In *IPTPS*, 2009.
- [12] R. Fuerderer, A. Herrmann, and G. Wuebker. *Optimal Bundling: Marketing Strategies for Improving Economic Performance*. Springer, 1999.
- [13] L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, and X. Zhang. Measurements, analysis, and modeling of bittorrent-like systems. In *ACM IMC*, 2005.
- [14] R. Joseph Lee and N. W. Alan. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988.
- [15] A. Legout, N. Liogkas, E. Kohler, and L. Zhang. Clustering and sharing incentives in bittorrent systems. In *ACM SIGMETRICS*, 2007.
- [16] A. Legout, G. Urvoy-Keller, and P. Michiardi. Rarest first and choke algorithms are enough. In *ACM IMC*, 2006.
- [17] N. Lev-tov, N. Carlsson, Z. Li, C. Williamson, and S. Zhang. Dynamic file-selection policies for bundling in bittorrent-like systems. In *IEEE IWQOS*, 2010.
- [18] D. S. Menasche, G. Neglia, D. Towsley, and S. Zilberstein. Strategic reasoning about bundling in swarming systems. In *IEEE GameNets*, 2009.
- [19] D. S. Menasche, A. A. Rocha, B. Li, D. Towsley, and A. Venkataramani. Content availability and bundling in swarming systems. In *ACM CoNEXT*, 2009.
- [20] R. S. Peterson and E. G. Sirer. Antfarm: efficient content distribution with managed swarms. In *USENIX NSDI*, 2009.
- [21] M. Piatek, T. Isdal, T. Anderson, A. Krishnamurthy, and A. Venkataramani. Do incentives build robustness in bittorrent. In *USENIX NSDI*, 2007.
- [22] M. Piatek, T. Isdal, A. Krishnamurthy, and T. Anderson. One hop reputations for peer to peer file sharing workloads. In *USENIX NSDI*, 2008.
- [23] E. W. Pugh. Origins of software bundling. *IEEE Annals of the History of Computing*, 24(1):57–58, 2002.
- [24] D. Qiu and R. Srikant. Modeling and performance analysis of bittorrent-like peer-to-peer networks. In *ACM SIGCOMM*, 2004.
- [25] P. Raghubir. Free gift with purchase: Promoting or discounting the brand? *Journal of Consumer Psychology*, 14(1-2):181–186, 2004.
- [26] M. A. Salinger. A graphical analysis of bundling. *Journal of Business*, 68(1):85–98, 1995.
- [27] A. Sherman, J. Nieh, and C. Stein. Fairtorrent: bringing fairness to peer-to-peer systems. In *ACM CoNEXT*, 2009.
- [28] M. Sirivianos, J. Han, P. Rex, and C. X. Yang. Free-riding in bittorrent networks with the large view exploit. In *IPTPS*, 2007.
- [29] Y. Tian, D. Wu, and K.-W. Ng. Analyzing multiple file downloading in bittorrent. In *IEEE ICCP*, 2006.
- [30] R. B. Wilson. Strategic models of entry deterrence. *Handbook of Game Theory with Economic Applications*, 1(1):305–329, 1992.
- [31] Y. Yang, A. L. H. Chow, and L. Golubchik. Multi-torrent: A performance study. In *IEEE MASCOTS*, 2008.
- [32] S. Zhang, N. Carlsson, D. Eager, Z. Li, and A. Mahanti. Towards a dynamic file bundling system for large-scale content distribution. In *IEEE MASCOTS*, 2011.