Characteristics analysis for the Cyberlockers Files

Study on Rapidgator Speedyshare and 1Fichier

Nan ZHAO
Télécom ParisTech
Paris, France
nan.zhao@telecom-paristech.fr

Loïc BAUD
DREV, Hadopi
Paris, France
loic.baud@hadopi.fr

Patrick BELLOT
Télécom ParisTech
Paris, France
bellot@telecom-paristech.fr

Abstract—**As the bandwidth of Internet rises from ISPs, the proportion of different Internet traffic and underling services have changed. There is about 20% decrease of the P2P traffic compared to an increase of more than 10% of the traffic of cyberlockers services, which are websites providing direct file download functionalities. In this paper we present a recent study over three cyberlockers: Rapidgator, Speedyshare and 1Fichier. Compared to prior studies, we applied a bias-free sampling method to randomly gather hosted files on the three cyberlockers. We aim at figuring out the characteristics of the hosted files on cyberlockers. In our work, we analysed and estimated the total number of files and the total size of files on these three cyberlockers. We specifically discussed the file size and file number distribution of hosted files in different file format and file content. Our results show that split-compressed files take a large part of the volume; there are many raw files on cyberlockers. Additionally, the results point that Rapidgator and 1Fichier are used to host entertainment and professional files while Speedyshare is mainly used to host private files.**

Keywords-cyberlocker; web mining; data analysis

I. INTRODUCTION

Cyberlockers are also referred as One-Click Hosting, which is a kind of website for file hosting and file sharing over the Internet. Cyberlockers allow Internet users to easily upload one or more files from users' local hardware devices such as computers, tablets, smartphones, etc. to a remote hosting server with only one click. In return, cyberlockers generate a URL for the uploaded file. The cyberlockers users can keep this URL for them or share it with their friends, and even can publish it on some websites such as forums.

From 2005 as the popularity of cyberlockers have increased [1], the proportions of Internet traffic has drastically changed. In Ipoque Internet traffic study 2007 [2] and 2008/2009 [3], they pointed out that the cyberlockers service traffic has increased at least 10% in Eastern Europe and Germany, while in South-western Europe it has increased more than 20%. Meanwhile, Peer-to-Peer traffic has a decrease of at least 10% from 2007 to 2008-2009. This shows that Internet users' preference is changing from the P2P to the One-Click Hosting service, such as cyberlockers. Compared to P2P service, cyberlockers service does not depend on the number of users who have the seeds. Once a file is uploaded on the server, it is available all the time until it is removed. Cyberlockers service is based on the HTTP protocol, or in some case FTP protocol, so it is easy to download files and there is no bandwidth limit for premium users. What is more, the IP addresses of uploaders and downloaders are only known by cyberlockers service. So it is difficult to trace and inspect the IP addresses of the devices that are connecting with cyberlockers sites [4]. All these conveniences explain why the Internet users preference changes, and why the number of cyberlockers increases.

Therefore, we would like to study over the cyberlockers service and figure out the characteristics of the content hosted on it. We used a statistic method to perform a random sampling of files from three cyberlockers Rapidgator, Speedyshare and 1Fichier. With retrieved information of each file, we analysed the general file size distribution, the files size distribution of different file content and file forms and the population of different content. We find that on cyberlockers there are more files with a small or medium size than those with a large size. However, they are long-tailed for the files larger than 1 GB. Additionally, Speedyshare is more for personal usage but also supports for the entertainment and professional usages. Rapidgator and 1Fichier are more for entertainment and professional usages. The highlights of our work could be summarised as below:

- The analysed files in our study are sampled directly from the three one-click hosting servers. Compared to tracing the user-end HTTP traffic in the prior studies, the sampled files in our study are more random, more representative and bias-free.
- Our study reveals that there are more than 18.7 millions accessible files on Rapidgator, more than 2.37 millions accessible files on Speedyshare and more than 2.32 millions accessible files on 1Fichier. The possible amounts of files size on Rapidgator, Speedyshare and 1Fichier are 6.93 PB, 0.18 PB and 1.0 PB separately.
- Both on Rapidgator and 1Fichier there are 60% of files no larger than 340 MB, which on Speedyshare is no larger than 40 MB. Compared to the other two, Speedyshare hosts most of small files. However, all the three cyberlockers are long-tailed for the files larger than 1 GB.
- More than 50% of files hosted on Rapidgator and 1Fichier are archive files, while on Speedyshare, the archive files only take about 38%. The split archive files hosted on Rapidgator and 1Fichier are about 37% and 39% separately, which can tell that there are a

large part of files hosted on the two cyberlockers to be accessible to the public. Additionally on all the three cyberlockers, the archive files represent more than 60% of the hosted volume.

- Both on Rapidgator and 1Fichier, there are more than 70% content of files can be classified as video. While on Speedyshare it is about 25%. Additionally on all the three cyberlockers, the file content of video takes more than 60% of hosted bytes.

The rest of the paper is organized as follows: the section II talks about the prior studies over the measurement and the traffic flows of the cyberlockers service. We describe our study method in detail in the section III. Then we analyse the sampled files in order to reveal the characteristics of cyberlockers in the section IV. In the last section V, we give the statements for this study and indicate the direction of our future work about the cyberlocker service.

II. Related Work

Cyberlockers turn to popular in the last ten years, so there are not so many studies over it yet. The most known one is Antoniades, Markatos and Dovrolis' paper [4]. In their work, they traced client-side traffic of a research centre and a campus networks. They show that cyberlockers traffic volume surpasses that of popular video service as YouTube and GoogleVideo. They also point out that the large objects as movie and software are often split into 100 MB or 200 MB files. They identify the traffic volume between Free and Premium users. There are more Free users than Premium users. In their work, they also measure that more than 70% of files are downloaded once and less than 0.05% files are downloaded more than five times.

Mahanti's work in [5] is based on monitoring the client-side traffic of a campus network. Besides this, [5] also crawled a cyberlockers search engine to get the published files. By observing the downloaded traffic size, this work states that most files are no larger than 100 MB. It also gives the proportions of different content on the cyberlockers. It shows that archive files take a large part on most of cyberlockers. In later work [6], Mahanti and his colleagues figure out the characteristics of cyberlocker traffic flows. This study focused on monitoring the TCP connections of a campus network. They find that more than 40% of campus hosts accessed to the top-10 cyberlockers. They also point out that although content flow traffic only takes 5% of total cyberlockers flows, it consume more than 99% of the traffic volume. They figure out that the cyberlockers flow size is about 3 MB, its flow duration is long-tailed and its arrival rate is much lower than HTTP's.

In Envisional' report [7], they crawled the web for URLs pointing to cyberlockers in order to estimate the percentage of copyrighted material exchanged via cyberlockers. They list that most of files hosted on cyberlockers are movies. The remaining of the content hosted on cyberlockers are music (10.1%), software (10%), game (9.4%), TV shows (8.5%), ebook (2.5%) and pornographic material (20%).

However, all the prior studies either monitored the client-side HTTP traffic, or crawled the forums, which publish files hosted on cyberlockers. This method is biased to the users' preference, which only can show what kind of files are popular to the cyberlockers users, but cannot represent the general characteristics of cyberlockers. Additionally, in [8] it mentioned the conception of Deep Web and Surface Web. Most of methods used in prior studies are biased to Surface Web. However according to [8], the size of Deep Web is 500 times larger than that of Surface Web, which means the prior studies are not enough representative and comprehensive.

The study in this paper is designed to avoid the bias caused by sampling over the Surface Web. It therefore:

- Takes into account the Deep Web;
- Is not based on the Internet monitor over a specific location in order to avoid the bias of users' behaviour.

III. Methodology

A. Sampling Methodology

Once user uploads a file on the cyberlocker, the server generates a URL for this file. This URL allows anyone to access the location where the file is hosted on the cyberlocker. On the Internet users' browser, it normally shows a webpage, which displays basic information of the file such as name, size and some optional information designed by the cyberlocker as file description, file upload date etc. The table below shows the URLs and its components on Rapidgator, Speedyshare and 1Fichier. From Table I we can tell that for Rapidgator, Speedyshare and 1Fichier, the number of possible ids are respectively 10^8 , $62^5 (= 9.2\times10^8)$ and $36^6 (= 2.2\times10^8)$ separately.

TABLE I. The Form of Cyberlockers URL

CLS	URL		
	URL Form	*Id L*	*Id elements*
Rapidgator	http://rapidgator.net/file/<id>	8	0-9[11]
Speedyshare	http://www.speedyshare.com/<id>	5	Alphanumeric
1Fichier	http://<id>.1fichier.com/	6	0-9, lowercase letters

Our sampling method is to generate random cyberlockers URLs. We use the random function of JAVA in order to randomly combine id elements to get a random URL. Our method can be summarised as following:

a) Generate a random cyberlocker URL;

b) Verify whether URL exists on the cyberlocker sever or not. If it does not exist, repeat a). If it does exist, continue to c);

c) Retrieve the file's information of file name and file size.

This method is based on large quantities of tests with different possible URLs of the three cyberlockers. This large quantity of tests can avoid the repeatability of the same URL. What is more, instead of crawling over the forum sites, it also

[1] We sampled Rapidgator on January 2013. At that time they still applied the old id form. Now they have changed with a new form made up by 30 letters and numbers.

can avoid the content bias caused by the users' preference and habit. This sampling process is running over network TOR with many different IP addresses located in different countries. The Rapidgator is sampled on January 2013, Speedyshare and 1Fichier are sampled on May 2013. In order to sample 1200 hosted files on each cyberlocker, we tested more than 13,000 possible URLs on Rapidgator, more than 590,000 URLs on Speedyshare and more than one million URLs on 1Fichier.

B. File Classification Methodology

In [5], they mixed file form and file content for the statistic study of different file types. In order to describe the different characteristics in content form and content type, we applied file form classification and file content classification respectively to analyse the sampled files. The file form is based on whether a file is a compressed file or not. The file content is based on the different content types. The following describes the two classifications.

File Form Classification

Single Archive Files: Compressed files that are not split.

Split Archive Files: Compressed files that are split.

Raw Files: Regular files.

File Content Classification

Audio: Files corresponding to music, concert and other audio record.

Document: Files corresponding to ebooks, magazines, all document formats and programming code.

Picture: Files corresponding to all image formats.

Software: Files corresponding to software, executable files and video games.

Video: Files corresponding to videos.

Others: Files that cannot set with any of the content types above.

For the second classification, in order to better understand the files content of each type and also in order to have a detailed distribution of each content type, we divide each content type into several sub-types. The Table II shows the detail of sub-types of each content type.

TABLE II. THE SUB-TYPES OF EACH CONTENT TYPE

Type	Sub-types Name
Audio	Music-Album Full/Part, Music-Song, Music-Clips, Music-Concert Live/Record, Others
Document	Book/Porn, Magazine/Porn, Articles, Simple Text, Code/Configuration
Software	Software Full/Part, Software Assistive, Video Game/Part, Video Game Assistive
Video	Film Full/Part, Film-Porn Full/Part, Film-Animation Full/Part, Series Full/Part, Series-Animation Full/Part, Media, Amateur, Tutor Full/Part, Others

IV. RESULT ANALYSIS

In this section, we analyse the sampled files on the three cyberlockers in order to find the general characteristics of cyberlockers and some specific properties for each cyberlocker.

A. Files Number and File Size

Firstly we would like to estimate the number of files hosted on each cyberlockers and the total hosted volume on each cyberlocker. We use the sampled file probability to estimate the total files number on a cyberlocker. The sampled file probability is the proportion of existing files over the total possible URLs, which in our study is the number of sampled files over the number of total tested URLs. Equation (1) is the formula to calculate the estimated number of files. N is the estimated number of hosted files, N_p is the number of all possible URLs, N_f is the number of sampled files, N_t is the number of tested URLs and N_f/N_t is the sampled file probability.

$$N = N_p \times N_f / N_t \quad (1)$$

As we mentioned in the chapter II, the possible number of URLs for Rapidgator, Speedyshare and 1Fichier is 10^8, 62^5 and 36^6 respectively. And the sampled file probabilities with all the tested URLs of the three cyberlockers are 18.73%, 0.26% and 0.11%. Therefore, we can estimate that there are more than 18.7 millions files hosted on Rapidgator, more than 2.37 millions files hosted on Speedyshare and more than 2.32 millions files hosted on 1Fichier.

However, there are some points that should draw attention:

- The same file can be posted several times by different users with different URLs. Therefore, there is content repeating in the hosted files on cyberlockers.
- There exist different versions for the same file content.
- The split archive files take a relative large part, which means that there are many content files for decompressing one useful file.

With the file size information on each cyberlocker, we can estimate the hosted files volume on cyberlockers. The average sizes for the sampled files are as below: Rapidgator 370 MB, Speedyshare 75.6 MB and 1Fichier 429 MB. Therefore, we can estimate the total volume by using average sizes multiple the estimated numbers of hosted files separately. Thence, the estimated volume of Rapidgator, Speedyshare and 1Fichier is 6.93 PB, 0.179 PB and 0.993 PB respectively.

B. File Size Distribution

In this section we analyse the file size distribution on the cyberlockers. Fig. 1a and Fig. 1b are the Cumulative Distribution Function (CDF) and Complementary CDF (C-CDF) of files size sampled from the three cyberlockers. From Fig. 1a we can see that on Speedyshare there is 49.95% of files whose size is no larger than 10 MB. This percentage is much higher than those of Rapidgator (3.17%) and 1Fichier (10.32%). On Rapidgator, we can find that the value of CDF increases fast for the file size no larger than 530 MB, which takes 82% of all Rapidgator files. And there are about 60% of files whose size is no larger than 330 MB. On Speedyshare, we find that its CDF increases largely for the size no larger than 140 MB, the rest part increases slowly. Additionally, we find that there are about 80% of files on Speedyshare that no larger than 140 MB. There are two inflection points on the CDF of

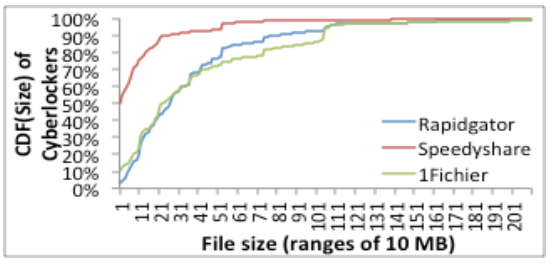
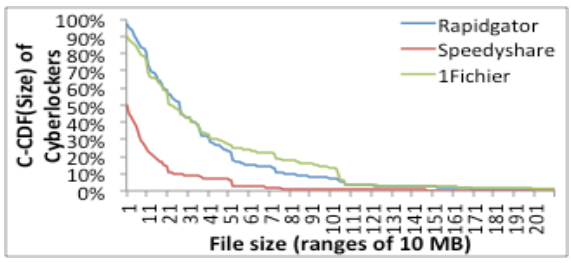


Figure 1. a.CDF of the files size on the three cyberlockers



b.C-CDF of the files size on the three cyberlockers

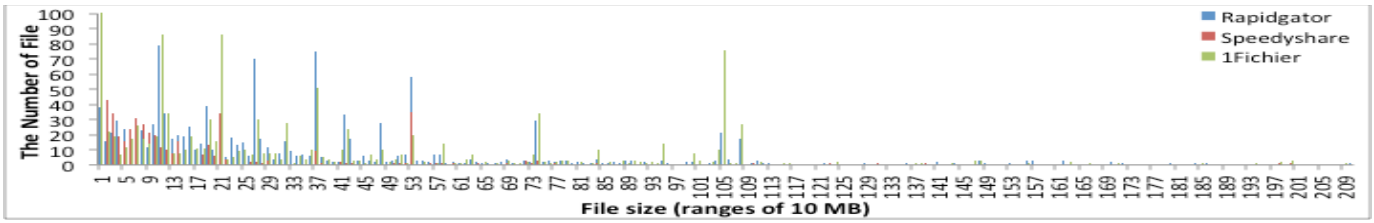


Figure 2. The distibution of the number of files with different size on the three cyberlockers

1Fichier. The first one is at 370 MB, where there are about 65% of files no larger than 370 MB on 1Fichier. The second one is at 1 GB, which shows that from 1 GB to 1.05 GB the augment value is about 8%.

As we mentioned in the last section that the average size of Rapidgator, Speedyshare and 1Fichier is 370 MB, 75.6 MB and 429 MB. With the CDF of the three cyberlockers, we firstly conclude that on Speedyshare mostly small files are hosted, which tends to a small average size. On Rapidgator, there are more than 30% of files between 300 MB and 1 GB, which lead to a medium average value. On 1Fichier there is also about 30% of files between 300 MB and 1 GB. However, for the files between 1 GB and 1.05 GB, the proportion is as high as 8%, which shows that there are many big files hosted on 1Fichier. That is why the average size on 1Fichier is larger than others.

Fig. 1b shows C-CDF of the three cyberlockers. We can see the proportion of size larger than 1 GB on Rapidgator, Speedyshare and 1Fichier is 7.5%, 0.97% and 13.57%. For the proportion of size larger than 2 GB, Rapidgator is 0.92%, and 1Fichier is 1.33%. Therefore, the files size on the three cyberlockers is long-tailed for the big files. Especially on 1Fichier, the files size is also heave-tailed.

Then we take a look at the peaks of the different file sizes on the three cyberlockers. Fig. 2 shows the distribution of the number of files with different sizes. On Rapidgator, we can see most file sizes drop before 530 MB. There are several peaks with a relative big number, which are at 10 MB, 110 MB, 120 MB, 190 MB, 270 MB, 370 MB, 420 MB and 530 MB. On Speedyshare, most file sizes drop before 210 MB, especially at the point of 10 MB, the peak value is 463[2]. The other peaks on Speedyshare are at 20 MB, 30 MB, 210 MB and 530 MB. On 1Fichier, most of file sizes drop before 740 MB. The peaks exist at 10 MB, 110 MB, 120 MB, 190 MB, 210 MB, 270 MB, 370 MB, 740 MB and 1.05 GB. We try to figure out what those peaks represent in the section C and D.

C. *Archive Files*

TABLE III. PROPORTIONS OF FILE NUMBER AND FILE SIZE IN FILE FORM CLIASSIFICATIO OF THE THREE CYBERLOCKERS

CLS	Single Archive Files		Split Archive Files		Raw Files	
	Number %	*Size%*	*Number %*	*Size%*	*Number %*	*Size%*
Rapid-gator	36.58	24.36	37.33	43.26	26.08	32.38
Speedy-share	28.17	32.92	9.67	38.01	62.17	29.07
1Fichier	15.40	10.54	38.55	50.10	46.04	39.36

In this part we analyse the files hosted on cyberlockers by applying the File form classification. Table III shows the file number and file size proportions of the three different file forms on Rapidgator, Speedyshare and 1Fichier,which can help us understand how they distribute on cyberlockers. On Rapidgator the proportion of single and split archive files are almost same, which is about 10% higher than that of raw files. On Speedyshare raw files take more than 60% of all files while split archive files take no larger than 10%. On 1Fichier raw files take more than 40% of all files, split archive files take about 38%, while the single archive files take the smallest part. We then take a look at the file size distribution. Among the three cyberlockers, we find that split archive files take the largest portion of the total file size. Both on Rapidgator and 1Ficher raw files take the second largest proportion of the total file size than that of single archive files. While on Speedyshare it is single archive files, which take the second large part of the total file size.

With the information shown in Table III, split archive files may not have the largest quantity but do have the biggest size portion. Also there are many raw files on the cyberlockers, which take 30% to 40% of the total file size. We infer that split archive files have larger size. Compared to [5], the archive files

[2] In order to better observe the file sizes, which do not have many files, we take 100 as the maximum of the y-axis.

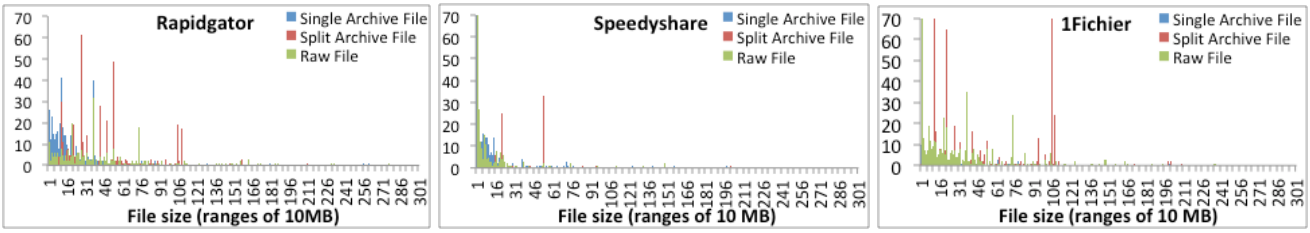


Figure 3. The distribution of the number of files with different sizes on the tree cyberlockers by file form classification

do not always have the largest quantity on cyberlockers.

In the second part we take a look at the distribution of the number of file with different size via file form classification. Fig. 3 shows this distribution on the three cyberlockers. Compared with Fig. 2, on Rapidgator the peak of 10 MB is from single archive files, the peaks of 110 MB and 120 MB are from single and split archive files. The peaks of 270 MB, 420 MB, and 530 MB are from split archive files. The peak of 370 MB is from single archive and raw files. On Speedyshare the peaks of 20 MB and 30 MB are from single archive and raw files. The peaks of 210 MB and 530 MB are from split archive files. On 1Fichier the peaks of 10 MB, 190 MB, 370 MB, 740 MB are from raw files. The peaks of 110 MB, 120 MB, 210 MB, 270 MB and 1.05 GB are from split archive files.

From this we can see that split archive files cause most of the peaks. While on 1Fichier raw files cause several obvious peaks as well. We also can infer that 110 MB, 120 MB, 210 MB, 270 MB and 500 MB could be the normal split archive files size.

D. File Content Analysis

TABLE IV. PROPORTIONS OF FILE NUMBER AND FILE SIZE IN FILE CONTENT CLIASSIFICATIO OF THE THREE CYBERLOCKERS

Content Type	Rapidgator		Speedyshare		1Fichier	
	Number %	*Size%*	*Number %*	*Size%*	*Number %*	*Size%*
Audio	20.42	9.28	17.83	14.47	3.83	1.15
Doc[3]	3.42	1.00	15.58	3.54	8.83	0.33
Others	3.33	2.62	6.50	7.33	15.24	16.81
Picture	1.83	0.15	24.50	0.67	1.50	0.04
Software	6.75	9.22	11.50	9.75	8.91	11.77
Video	64.25	77.73	24.08	64.24	61.70	69.90

In this section we analyse the files hosted on cyberlockers by file content classification. Table IV shows the number and the size proportions for the files of different content types. We can find that both on Rapidgator and 1Fichier the video files take more than 60% of all the files. On Speedyshare the proportion of the number of files is quite evenly distributed among the categories. And compared to the other two cyberlockers, picture files take a relative lager proportion of 24% on Speedyshare. For the file size proportion, video files take the largest part on all the three cyberlockers. Even on Speedyshare, the 24% files take 64% of the total size. Additionally, software files take about 10% of the total size on all the three cyberlockers. Compared to the other two, audio files just take a very small part of the total size on 1Fichier.

Then we take a look at the distribution of the number of files with different size via file content classification. Fig. 4 shows this distribution on the three cyberlockers. Compared with Fig. 2, we can figure out which content types cause peaks in it. Firstly, we can see that video files cause most of peaks on the three cyberlockers. On Rapidgator the peak of 10 MB is from document and picture files. The peaks of 110 MB and 120 MB are from audio and video files. The remaining peaks are all from video files. On Speedyshare all the types of files have a large quantity for the peak of 10 MB. The peak of 20 MB is from audio files and the remaining peaks are all from video files. On 1Fichier the peak of 10 MB is from document files and all the remaining peaks are from video files.

Compared with the file form classification result, on Rapidgator, at 270 MB, 420 MB and 530 MB video files are in the form of split archive and at 370 MB video files are single files in the form of archive or raw. On Speedyshare at 20 MB and 30 MB, audio and video files are mostly in the form of split archive. The video files of 210 MB at 530 MB are in the form of split archive. On 1Fichier the peak of 10MB are document files, which are raw files. The video files at 190 MB, 370 MB and 740 MB are mostly in the form of raw. The video files at 110 MB, 120 MB, 210 MB, 270 MB and 1.05 GB are mostly in the form of split archive. Therefore we can infer that video files are most divided into split-compressed files of 210 MB, 270 MB and 500 MB. And video files of 370 MB or 740 MB are mostly entire videos without compressing.

Then we take a look at the file number and file size distribution in sub-content types. Fig. 5 shows the general file number and file size distributions of each sub-type. On Rapidgator the files of music album, full and part porn film and series take obvious proportions of the total file number. For the file size distribution, the files of full and part film, full and part porn film and series have the most parts of the total file size. On Speedyshare files of picture take about 24% of all the total file number. While files of music album, part film, part video tutor and series take the most part of the total size. Especially for the part video tutor files, they are only 3% of the total file number but take more than 18% of the total size. On 1Fichier the files of part film, series and series animation have evident high proportion of the total file number. And files of part video game, full and part film and series take the most part of the total size. On 1Fichier the type content of others takes a relative large part both in file number and file size distribution. That means there is a large proportion of personal files hosted

[3] Doc here represents the content type Document

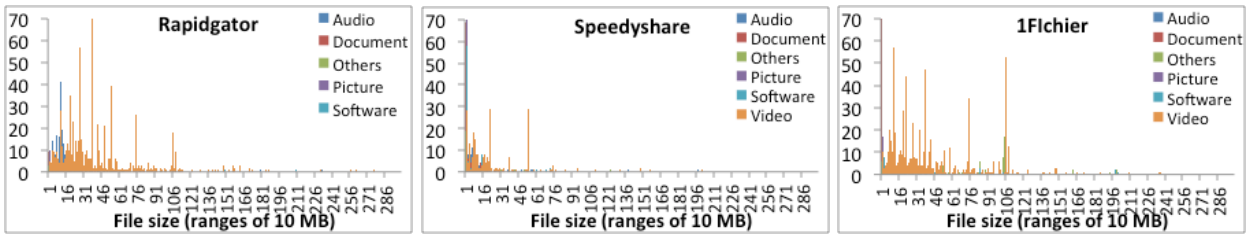


Figure 4. The distribution of the number of files with different sizes on the tree cyberlockers by File form classification

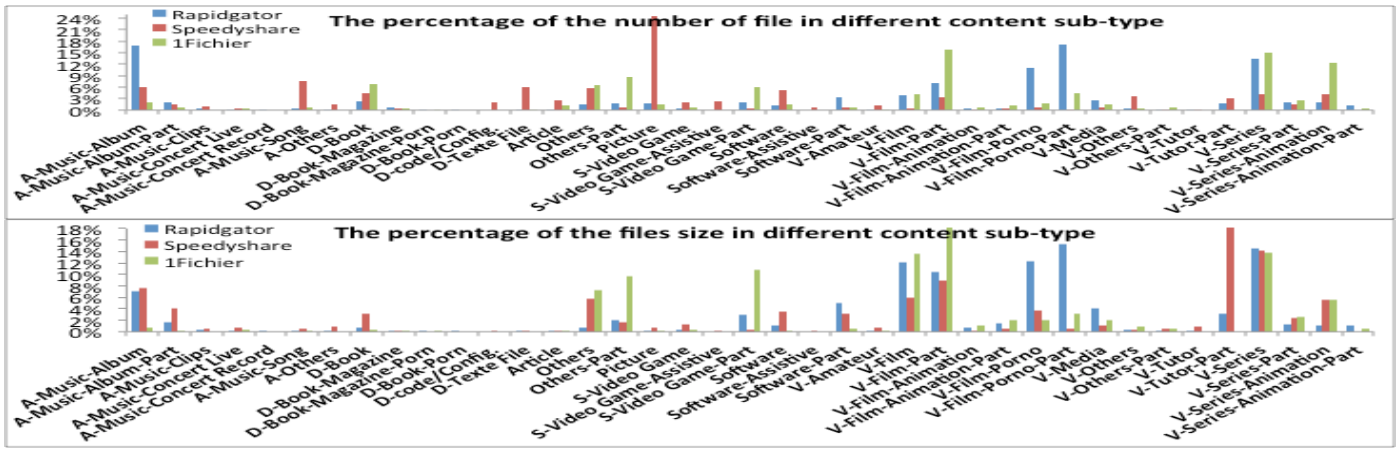


Figure 5. The file number and file size distribution of files with sub-content typs on the three cyberlockers

on 1Fichier. Compared to [7], we find that film files do not always have the largest quantity on cyberlocker. From the number and size distribution of sub-content types, we also can tell that Rapidgator and 1Fichier are mainly for entertainment and professional usages. And on Speedyshare there is a relative high proportion of files for the personal usage.

V. Conclusion and Future Work

In this paper we present our recent work about the characteristics of the hosted files on cyberlockers. We applied a statistic method to study files hosted on Rapidgator, Speedyshare and 1Fichier. We estimated the total file number and the total file size on the three cyberlockers. There are more than 2 millions files hosted on Speedyshare and 1Fichier and 18 millions on Rapidgator. The possible total size is 6 PB on Rapidgator, 0.18 PB on Speedyshare and 1 PB on 1Fichier. Then we focused on analysing the file size and file number distribution in different file form and file content type. We find that the split-compressed files do not take the largest portion of the total file number, while they do really take the largest portion of the total volume of all the three cyberlockers. There are many raw files hosted on the cyberlockers. We also find that video files around 200 MB and 500 MB mostly are split archived files; video files around 300 MB and 700 MB mostly are raw files. Additionally, video files take the largest part of the total size among all the content types. There are many music, films, TV series and animation series on Rapidgator and 1Fichier, which also take a high proportion of the total file size on the three cyberlockers. With the content type study we can tell that Rapidgator and 1Ficher are most for the entertainment and professional usage that the files hosted on them are always the published files on websites for sharing and downloading. And for Speedyshare, there is a great part for the personal usage to store the personal files.

For the future work, first we will continue to study the detailed characteristics of the files of the video type. Then we would like to work on the lifecycle of the hosted files on cyberlockers. In the last part, we would like to figure out the relation between the cyberlockers and the file-sharing and file-downloading forums.

References

[1] File hosting service, http://en.wikipedia.org/wiki/File_hosting_service #cite_note-2.

[2] Hendrik Schulze and Klaus Mochalski, “Internet study 2007,” Ipoque, 2007.

[3] Hendrik Schulze and Klaus Mochalski, “Internet study 2008/2009,” Ipoque, 2009.

[4] Demetris Antoniades, Evangelos P. Markatos and Constantine Dovrolis, “One-Clike hosting services: a file-sharing hideout,” IMC’9 Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, 2009, pp. 223–234.

[5] Aniket Mahanti, “Measurment annd analysis of cyberlocker services,” WWW’11 Proceedings of the 20th international conference companion on World wide web, 2011, pp. 373–378.

[6] Aniket Mahanti, Niklas Carlsson, Martin F. Arlitt and Carey Williamson, “Characterieing cyberlocker traffic flows”, IEEE 2012 37th conference on Local Computer Networks (LCN), 2012, pp. 410–418.

[7] Envisional, “An estimate of infringing use of the Internet,” Technical Report, 2011.

[8] Micheal K.Bergman, “The deep web: surfaceing hideout value,” Journal of Electronic, Vol.7, No.1, 2001.