# One-Click Hosting Services: A File-Sharing Hideout

Demetris Antoniades
FORTH-ICS
Heraklion, Greece
danton@ics.forth.gr

Evangelos P. Markatos
FORTH-ICS
Heraklion, Greece
markatos@ics.forth.gr

Constantine Dovrolis
GATECH
Atlanta, USA
dovrolis@cc.gatech.edu

## ABSTRACT

File sharing using peer-to-peer (p2p) systems is a major Internet application and the leading source of network traffic today. However, the dominance of p2p systems for file sharing has been recently challenged by an increasing number of services, such as RapidShare and MegaUpload, which offer users the ability to share files through centralized servers, without relying on an underlying p2p infrastructure. These services, referred to as One-Click Hosting (OCH), have the potential to offer users better performance and availability than p2p systems. If they succeed, OCH services may become the leading platform for file sharing and eventually replace p2p systems for this purpose. In this paper, we present the first, to our knowledge, detailed study of OCH traffic and services focusing on the most popular such service: RapidShare. Through a combination of passive and active measurements, we attempt to understand their service architecture, usage patterns, and content characteristics. We also compare RapidShare with BitTorrent in terms of user-perceived throughput and content availability, and we explore the characteristics of some popular RapidShare indexing sites.

## Categories and Subject Descriptors

C.2.0 [**Computer Communication Networks**]: General; H.3.5 [**Information Storage and Retrieval**]: Online Information Services,Web based services

## General Terms

Measurement, Performance

## Keywords

One Click Hosting, File Sharing, Peer-to-Peer, BitTorrent, RapidShare

## 1. INTRODUCTION

Over the past decade, file sharing has become one of the most popular Internet user activities, surpassing in terms of traffic volume email, ftp, and even the mighty World Wide Web. Indeed, file sharing accounts for the largest portion of network traffic, reaching as high as 60-70% of the total volume in some ISPs [15]. This popularity of file sharing was fueled by the evolution of the p2p paradigm, which enabled users to exchange files without having to rely on a third-party server infrastructure.

In the last couple of years, however, a new type of file sharing service has emerged, usually referred to as One-Click Hosting (OCH). OCH sites, such as RapidShare, MegaUpload or FileFactory, allow users to share files through dedicated server infrastructures. Using OCH services, a user can share a large file (even gigabytes) with one or more other users in a sequence of few simple steps: (i) The user uploads the (potentially encrypted) file on an OCH server through a basic Web interface. (ii) The OCH service provides the uploader with a URL for the file. (iii) The uploader shares that URL (and a decryption key, if necessary) with other users either privately (e.g. email) or through public indexing sites, such as rslinks.com, blogs, personal Web pages, etc. (iv) For publicly indexed content, users can find the download-URL through search engines, and they can then download the file through a basic Web interface. Usually, OCH services offer both Free and Premium (i.e. paying) accounts as well as several incentives (discussed later) for attracting users to upload content (making the service more valuable) and to subscribe as Premium users.

Compared to p2p applications, OCH services could provide several advantages for file sharing:

- **Availability:** Files uploaded on OCH services are available 24/7 - not only when a "seeder" is available as is the case with p2p systems.
- **Anonymity:** In the p2p paradigm, peers have to disclose their IP addresses to the community when they are downloading (or uploading) a file. Thus, someone can infiltrate a p2p system and harvest IP addresses of hosts that engage in file sharing. On the contrary, in OCH services the IP addresses (and in some cases the identity) of the uploaders and downloaders are known only to the OCH service. Unless the OCH service cooperates, it would be difficult for someone to harvest IP addresses of computers that engage in file sharing through OCH sites.
- **Performance:** Because of their business model and always-on infrastructure, OCH services aim at offering high throughput, at least to their Premium users. Indeed, our experiments in Section 7.1 show that a Premium RapidShare user receives an order of magnitude higher throughput than a BitTorrent user downloading the same file.
- **Content Availability:** As we show in Section 7.2, even though a user can find popular content in both OCH and p2p ser-


*IMC'09,* November 4–6, 2009, Chicago, Illinois, USA.

vices, less popular/known content can be found more frequently at OCH services.

- **Incentives:** OCH services allow frequent uploaders to receive a higher download throughput, providing the incentive to upload more files, thus making OCH services richer in content.

In this paper, we study the OCH paradigm from multiple perspectives, focusing on RapidShare, the most popular (at least in terms of traffic) OCH service today [15]. Through a combination of passive and active measurements, we analyze traffic characteristics, usage patterns, and especially, we explore their service in terms of location, multihoming, load balancing, file allocation policy, etc. Further, we compare RapidShare and BitTorrent in terms of user-perceived performance and content availability. Finally, we look at the major OCH indexing web sites, aiming to characterize the type of content users share and whether that content is copyrighted.

The contributions of this paper are:

1. We provide the *first* to our knowledge study of OCH services focusing on their client, server, traffic, content and performance characteristics. This study improves our understanding of OCH services and it allows a comparison with other similar services, such as Web Content Distribution Networks (CDN).

2. Using passive network monitoring, we measure the volume of OCH traffic and show that it surpasses (or it is comparable to) the traffic volume of popular video services such as YouTube and GoogleVideo.

3. We show that most files are requested only once in our client-side traces (collected at two academic networks) - very few files were requested more than five times during our 5-month long traces. This suggests that caching, often used in CDNs, may not be effective in OCH services.

4. We explore the number of users that upload content, based on indexing sites, and show that it is only a handful of users that provide most OCH content, making the entire system quite sensitive to minor changes in the number and "productivity" of uploaders.

5. By using the Tor anonymity network as a large geographically distributed client base, we develop a methodology to infer the server location, content replication approach and load balancing that RapidShare uses.

6. We compare the relative benefits of OCH and p2p file sharing systems using RapidShare and BitTorrent, in terms of performance and content availability. We show that RapidShare provides significant benefits in these two aspects.

7. Our analysis of few public indexing sites shows that users often rely on OCH services to share movies, songs, games, and software, and that in most cases this content is protected by copyright legal constraints making such file sharing illegal.

The rest of the paper is organized as follows: We introduce OCH services in Section 2 and describe our datasets in Section 3. Based on the collected data, we provide a characterization of OCH traffic during our monitoring period in Section 4, and a characterization of OCH clients in Section 5. We attempt to infer the architecture of RapidShare (the leading OCH service) in Section 6. We compare user experience between RapidShare and BitTorrent (the leading p2p system) in Section 7. Section 8 explores the characteristics of some popular OCH indexing services. Finally, we place our work in the appropriate context by presenting the related work in Section 9 and we conclude the paper with a summary of our main findings in Section 10.

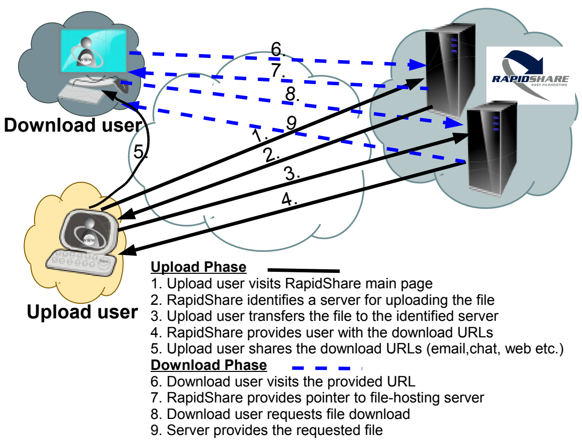



**Figure 1: Sequence of steps to share a file through RapidShare.**

## 2. ONE-CLICK HOSTING SERVICES

Since early 2005 several OCH sites made their appearance, including megaupload.com, rapidshare.com, filefactory.com, and others. These sites facilitated the creation of a vibrant user community that shares files reliably and inexpensively. OCH services can be described as Web services that allow a user to upload and store files on dedicated, always-on servers, and then share those files with other users through a URL (see Figure 1). The whole process is usually offered for free and needs a small number of steps for both uploading and downloading files. OCH services do not offer indexing or search capabilities, and search engines cannot crawl the download URLs. What happens, however, is that there are other "indexing" Web sites, such as EgyDown.com or RapidShareIndex.com, in which upload users can post OCH URLs for the files they want to share. Those indexing sites are often publicly visible and crawled by search engines.

In this paper we focus on the largest and most popular OCH service: RapidShare. RapidShare was launched in October 2006, and within two years it grew to 2.5 million users.[1] In our client-side traffic traces (§ 3), RapidShare is responsible for more than 50% (on average) and up to 95% of the total traffic exchanged through OCH Services. [2]

The main page of RapidShare offers a basic Web interface containing just the necessary fields to upload a file. Once a file has been uploaded, RapidShare provides the user with two URLs: the "download URL" that is to be shared with other users who want to access the file, and the "remove URL" to request deletion of that file. To create a viable business model, RapidShare offers two types of service: "Free" and "Premium". **Premium users** can enjoy unlimited use of RapidShare resources: their upload and download bandwidth is not throttled, they can start several concurrent downloads, upload files as large as 2GB, etc. On the other hand, **Free**

[1] http://siteanalytics.compete.com

[2] Similar results are shown by ipoque's Internet studies [14].

**users** are given a download bandwidth that is no more than 200-2000kbps, they can do only one concurrent download at a time, the maximum size of an uploaded file is no more than 200MB (it used to be 100MB until recently), [3] there is a mandatory wait time between successive downloads, etc. To entice users to upload as much content as possible, RapidShare offers "points" to an uploader each time her content is downloaded. Points, can, in turn, be exchanged for Premium accounts or extra download capacity. Such incentives have quickly enabled RapidShare to store a huge variety of songs, movies, games, software, books and other types of content. Last but not least, it is important to note that RapidShare is not responsible for any copyright violations due to illegal file sharing conducted using their infrastructure. Uploaders need to declare that the uploaded content is not protected by copyright laws, and RapidShare is further legally protected by not offering any indexing or search facilities for the content they store.

## 3. MAIN DATASETS

Table 1 summarizes the main dataset of this study - we refer to this dataset as our "client-side" traces. We collected flow traces in the IPFix format, as well as HTTP packet headers, at two monitoring locations: `Monitor1` is the main Internet access link of a National Research Network that serves a population of about 10,000 students and academics. `Monitor2` is the main Internet access link of a University campus network of about 1,000 students and faculty. Both monitoring points are located in Europe and have a user base composed mostly of university students. One may argue that these demographics are biased and not general in terms of age and location. Our analysis, however, focuses on information that would not depend, most likely, to these demographics. Data from Monitor1 cover about 4 months, while the Monitor2 data cover about 5 months. Table 1 also mentions the total number of unique client IP addresses that access RapidShare during the monitoring period from each of our monitoring sites. We saw about 750 IP addresses at Monitor1 and 450 IP addresses at Monitor2. These numbers should be viewed as an upper bound on the number of actual RapidShare users in the two monitored networks because some users may be using DHCP. The identification of RapidShare flows was performed using the HOST header field of HTTP requests (searching for the string "rapishare.com"). Based on that data, we could then identify all relevant RapidShare flows in our traces. We also use some additional datasets that are described later in the paper, when first introduced.

## 4. OCH TRAFFIC VOLUME

In this section we examine the traffic volume of OCH services, and of RapidShare in particular, in our client-side traces. We compare the traffic volume that these services generate with Web and BitTorrent traffic, as well as with major video streaming services.

As shown in Table 1, in Monitor1 OCH services generate 3.32% of the total traffic volume. This is low compared to BitTorrent traffic (44.5%) but a significant fraction (25%) of the total HTTP traffic (12.8%). The fraction of OCH traffic is significantly lower in Monitor2. Figure 2 shows the aggregate traffic rate downloaded from all OCH services in hourly intervals (the curve for Monitor2 includes 5 more months of data). [4] The hourly rates vary widely and, even though the long-term average is only about 1Mbps in either monitor, the hourly OCH rate often reaches up to 10-20Mbps, which

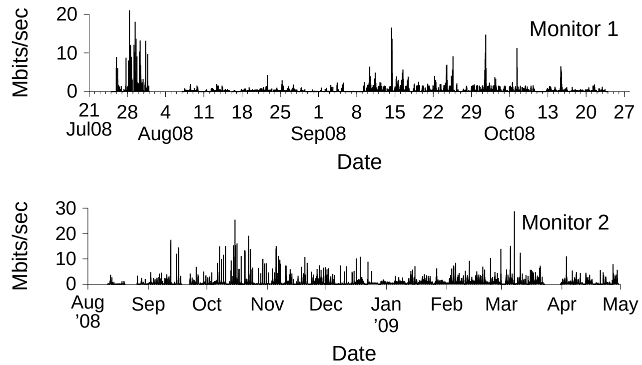



**Figure 2: Download traffic rate (at the two monitored sites) from all major OCH services in hourly intervals.**

is significant compared to the total traffic rate in these traces. We were hoping to see a clear trend in this timeseries, but this is not the case even in the 9-month period covered in Monitor2.

RapidShare generates more than 80% of the OCH traffic volume in our traces; so, in the rest of this paper we focus on this particular OCH service. Figure 3 compares the average daily download rate of RapidShare with two popular video streaming sites: YouTube (`www.youtube.com`) and GoogleVideo (`googlevideo.com`). Note that RapidShare generates more (Monitor1) or comparable (Monitor2) traffic volume than these major content providers.

Finally, Figure 4 shows the hourly traffic rate for Web and OCH traffic during a randomly selected week in September 2008 at Monitor1. Note that OCH traffic follows a similar diurnal pattern with Web traffic: much less activity during the evening hours and weekends. Afterall, OCH can be viewed as just another Web service.

## 5. CHARACTERIZATION OF RAPIDSHARE CLIENTS

In this section, we focus on the characteristics of RapidShare clients that are active in our two university traces. As mentioned in Table 1, an upper bound for the number of RapidShare clients during the course of our study is around 750 for Monitor1 and 450 for Monitor2.

### 5.1 Flow sizes and client downloads

First, we analyze the flow sizes downloaded by RapidShare clients. Figure 5 shows the CDF of the download flow sizes for all RapidShare connections. 50% of the flows are smaller than 700B at Monitor1 and 7KB at Monitor2, while 90% of the flows are smaller than 150KB at both monitors. Since most files provided by RapidShare are typically several megabytes long, these smaller flows probably correspond to Web page accesses, failed/stopped downloads or Web page refreshes to see the remaining wait time until the next download can start. The remaining 10% of the flows, which transfer more than 150KB, are probably actual downloads from RapidShare. In the rest of the paper we use a threshold of 150KB to distinguish between "content download flows" (larger) and "browsing flows" (smaller). As shown in Figure 5, the CDF increases slowly after 100KB or so, meaning that the identification of content download flows should be robust to the selection of this threshold.

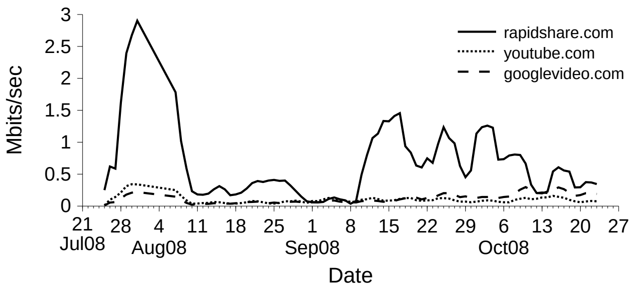
Figure 6 shows the complementary CDF (C-CDF) of content download flow sizes. Up to the point of 100MB, flow sizes appear to be Pareto distributed, as probably expected. For larger flows we

[3]To overcome this file size limit, uploaders partition large objects in files of that size and share them through several URLs.

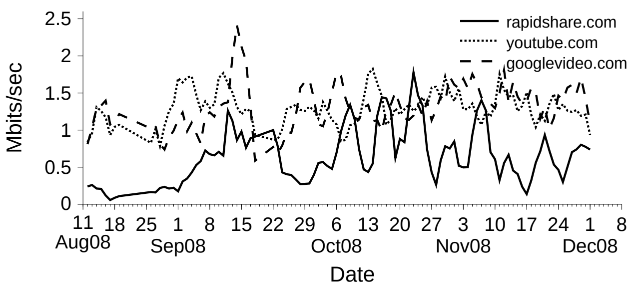
[4]We used the list of OCH services given at `http://en.wikipedia.org/wiki/One-click_hosting`.

| Name | Collection period | Tot. Bytes | Flows | HTTP | BitTorrent | OCH | RapidShare | RapidShare Clients |
|---|---|---|---|---|---|---|---|---|
| Monitor1 | Jun 6 - Oct 23'08 | 60.8TB | 2.2B | 12.8% | 44.5% | 3.32% | 2.7% | 748 |
| Monitor2 | Aug 10 - Dec 2'08 | 214.8TB | 1.4B | 4.72% | 56.4% | 0.23% | 0.22% | 449 |

**Table 1: Description of the client-side datasets.**




(a) Monitor 1




(b) Monitor 2

**Figure 3: Download traffic rate from RapidShare, YouTube and GoogleVideos in daily intervals.**

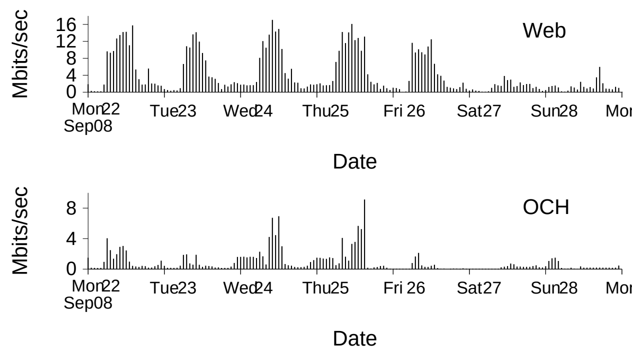



**Figure 4: Download traffic rate for HTTP and RapidShare in hourly intervals during a week in September'08 (Monitor1).**

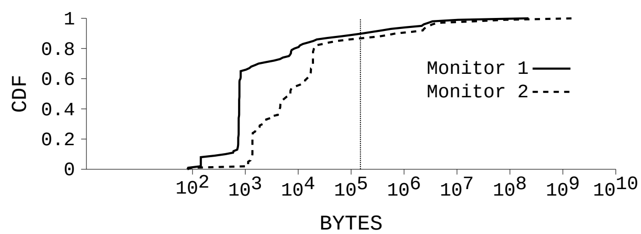



**Figure 5: CDF of RapidShare download flow sizes at the two monitors.**

note two significant drops, one at 100MB and another at 200MB. These sizes correspond to the maximum upload file size limits that RapidShare enforces for free uploaders: this limit was 100MB and was increased to 200MB in October'08. Premium users, on the other hand, can download and upload files up to 2GB. The difference between the two distributions for files larger than 200MB implies that there are much fewer Premium users at Monitor1 than at Monitor2.

Next, we examine the number of daily content downloads per user (or client). We assume that a client uses the same IP address during the day, and so the daily downloads from the same IP address are interpreted as downloads from the same user. Most clients perform more than one download per day (57% of clients at Monitor2), and only 23% of clients perform more than 10 downloads in the same day. Note that large objects (movies, software, etc) are often split into 100MB or 200MB files, and so a large number of consecutive flows by the same user may correspond to downloading different parts of the same object. Figure 7 shows the C-CDF of the daily number of content download flows per client at Monitor2. This empirical distribution can be approximated by a Pareto distribution with shape parameter 0.66. This low value of the shape parameter implies extremely large variability, to the point that neither the variance nor the mean of the underlying distribution are well-defined. The C-CDF for Monitor1 is similar.

## 5.2 Premium vs. Free users

As previously mentioned, RapidShare supports two user types: **Free** and **Premium**. All content is available to both types and the main difference is that the former are limited in terms of their upload flow sizes, concurrent downloads, and download throughput. RapidShare reports that the download throughput of Free users is throttled to 200-2000kbps. In this section we attempt to identify Premium users based on their download throughput, assuming that content download flows that receive more than 2Mbps are generated by Premium users. Of course, we may underestimate the number of such users when their throughput is limited by their access link capacity or by Internet congestion, and not by RapidShare.

Figure 8 shows the distribution of the average download rate per user observed on a daily basis. The rates were calculated considering only content downloads flows. Almost all users at Monitor1 ( 95%) and more than 80% of users at Monitor2 experience an average throughput that is less than 2Mbps, suggesting that most of them are Free users. At Monitor2, the remaining 20% (5% for Monitor1) of the users enjoy throughput up to 14Mbps and they must be Premium customers. Though, the percentage of Premium users is not very large, and may vary based on the user demographics and geographic location, it is important to note that there exists a significant fraction of users willing to pay a fee to use a service that is also offered for free, as long as they can enjoy some "premium features".[5]

[5]The artful reader may argue that cracked or phished Premium accounts are sometimes available. On the other hand, RapidShare can detect and block these accounts as easily as a typical user can locate and use them.

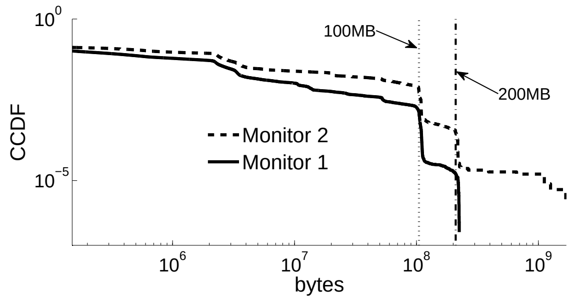



**Figure 6: C-CDF of RapidShare content download flow sizes at the two monitors.**

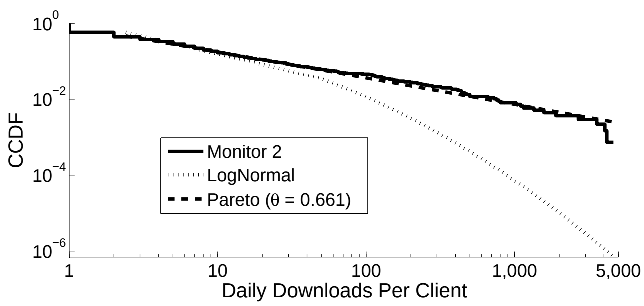



**Figure 7: Distribution of daily content downloads per client at Monitor2.**

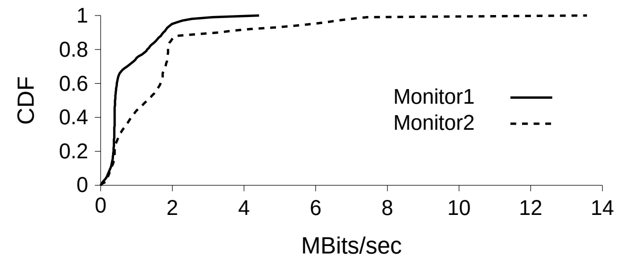



**Figure 8: Distribution of the average content download throughput per user, measured on a daily basis.**

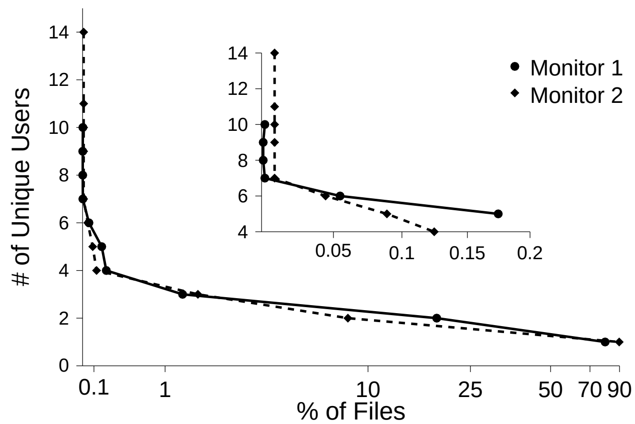



**Figure 9: File popularity measured by the number of clients that downloaded each file.**

## 5.3 File popularity

Here, we examine the files shared by RapidShare clients in our client-side traces. Due to privacy concerns, we limit our analysis to the number of unique files, ignoring the actual filenames. We further examine the type of content shared using RapidShare in Section 8. Note that the same content object (a movie or a song, for instance) can be stored as several different files in RapidShare (aliases). The following analysis focuses on the popularity of individual files, not of the underlying content. We should first note that we only focus on file popularity as seen at our two monitoring sites. Obviously, we cannot make any statements about the popularity distribution of different files in a wider scale. Our main focus is to examine whether caching RapidShare content close to clients would make sense or not.

Figure 9 shows the popularity of each file as the number of clients that downloaded that file in our traces. More than 75% of the files were downloaded only once. The inner plot focuses on the distribution of the most popular files. Very few files are highly popular among the clients of each monitor; less than 0.05% of the files were downloaded more than five times, and only a handful of files were downloaded more than ten times. These results suggest that there would be little benefit to cache RapidShare files close to clients, arguing in favor of a centralized infrastructure where all servers reside at the same location. Indeed, as the next section shows, this appears to be the case with RapidShare. Also, this file popularity distribution makes RapidShare very different than traditional CDNs that rely heavily on caching popular Web objects close to clients and that maintain cache hit rates of more than 90% [2, 12].

## 5.4 Summary

This section performed a characterization of RapidShare client behavior using client-side traces from two university networks. Our results show that more than half of the clients perform more than one file download per day. These downloads are mostly performed by non-paying users, who experience download throughput up to 2Mbps. However, a significant fraction of users (around 12%) is willing to pay a small fee to get a better service. The premium users seen in our client-side traces experience download throughput up to 14Mbps (i.e. downloading a 200MB file in less than 2 minutes). The users' daily activity mainly involves downloading a small number of files, which often corresponds to a single media object.

In terms of traffic patterns, even though RapidShare can be viewed as just another Web service, the popularity of unique files downloaded by its clients differs significantly from that of traditional Web browsing. We observed only a small number of files being downloaded more than once during the whole monitoring period. This suggests that caching RapidShare content close to clients would offer little or no benefit.

# 6. SERVICE ARCHITECTURE

In this section we attempt to understand the RapidShare architecture based on information from our client-side traces as well as from active measurements. In particular, we explore the number of deployed servers, their network connectivity, geographical location, load balancing and content replication strategies.

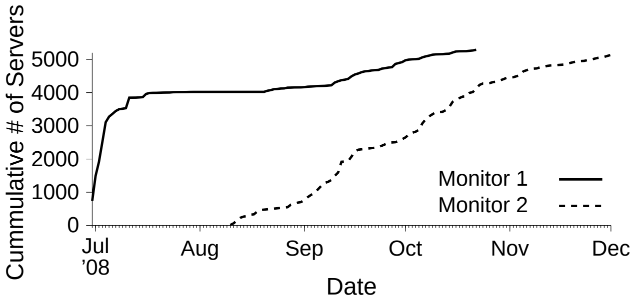



**Figure 10: Cumulative number of RapidShare server IP addresses seen in our traces.**

## 6.1 Number of servers

Figure 10 shows the cumulative number of RapidShare server IP addresses seen in our client-side traces. Recall that we identify such servers from the HOST header field of the HTTP requests seen at the two monitors. We observed 5,291 RapidShare server IP addresses at Monitor1 and 5,135 server IP addresses at Monitor2. The slight difference should be expected given there are more clients, and thus more sessions, at Monitor1. The increase in the number of servers in early September is interesting: RapidShare had announced that they will increase their server infrastructure at the same time period. [6]

Note that each IP address does not necessarily correspond to a distinct host. It is possible that the same host has several network interfaces, or that the same physical host is used as multiple virtual servers with distinct IP addresses. We attempted to infer, using the IP-ID method [3], whether different server IP addresses generate packets with interleaved IP-ID values, but we could not find any. Of course, this test does not exclude the possibility of server virtualization, as two virtual servers on the same physical host would run different IP stacks.

## 6.2 Address blocks and upstream ISPs

At both monitors, we observed that the server IP addresses belong to 36 distinct /24 subnets (the term "subnet" refers to a /24 prefix block in this paper). We looked up the origin-AS of these subnets and the results are shown in Table 2. The 36 subnets are allocated to 8 ISPs. This large degree of multihoming is typical for large content providers, such as RapidShare [8]. Multihoming can improve the reliability, performance and transit costs of a content provider. In particular, a content provider would prefer to balance its load among upstream providers so that the 95-th percentiles of its outgoing traffic through each provider remain as low as possible. There are commercial "intelligent route control" systems that perform such load balancing optimizations [11].

## 6.3 Server locations

With such a large number of servers and upstream providers, we may expect that RapidShare deploys servers in a large number of different geographical locations, similar to standard CDN practices. To explore this issue we probed the observed RapidShare server IP addresses from multiple geographical locations (landmarks) using `traceroute`. Our landmarks were several Planetlab hosts in different countries around the globe [6,28]. Our geolocation method is simple and it is based on the minimum Round Trip Time (RTT) between each landmark and a server. Katz-Basset *et al.* [19] showed that shortest RTT measurements using `ping` can provide geolocation results of comparable accuracy to more complex methods. An alternative and simpler method would be to use a geolocation database to query for the server's location, though some initial experiments with such a database gave as inaccurate results mapping each IP address to the location of the corresponding ISP.

[6]http://rapidshare.com/news.html

| ISP | AS | Subnets |
|---|---|---|
| Level 3 | 3356 | 195.122.131/24, 195.122.149/24, 195.122.151/24, 195.122.152/24, 195.122.153/24, 212.162.63/24, 62.140.31/24, 62.67.46/24, 62.67.50/24, 62.67.57/24 |
| GlobeInternet | 6453 | 195.219.1/24, 80.231.128/24, 80.231.24/24, 80.231.41/24, 80.231.56/24 |
| GBLX | 3549 | 206.57.14/24, 208.48.186/24, 64.211.146/24, 64.214.225/24, 64.215.245/24 |
| Fidelity | 22958 | 207.138.168/24 |
| INETBONE | 25074 | 212.162.2/24 |
| DTAG | 3320 | 217.243.210/24, 62.153.244/24, 80.152.63/24 |
| Cogent | 174 | 82.129.33/24, 82.129.35/24, 82.129.36/24, 82.129.39/24 |
| TeliaNet | 1299 | 80.239.137/24, 80.239.151/24, 80.239.152/24, 80.239.159/24, 80.239.226/24, 80.239.236/24, 80.239.239/24 |

**Table 2: Address blocks and transit providers used by RapidShare .**

Figure 11 shows the RTT results. The landmark locations are shown in the x-axis. Each point in this plot is the minimum RTT measurement between the corresponding landmark and the servers of a RapidShare subnet. There are 36 points for each landmark, one for each subnet. We sorted the landmarks according to the minimum measured RTT across all subnets. Landmarks with RTTs lower than 100ms are located in European countries. The lowest RTTs come from landmarks in central Europe (Netherlands, Germany and France). Landmarks in the US, Brazil or Japan give much higher RTTs for all subnets. In some cases, the subnet measurements from the same landmark are grouped in clusters; each cluster corresponds to a different routing path from the landmark to the corresponding subnets. The results of Figure 11 suggest that all RapidShare subnets are probably located somewhere in central Europe.

To identify the location of servers more accurately, we also examined the 2-3 last hops returned from `traceroute` towards all RapidShare IP addresses, from a single landmark. There were 48 distinct names for the penultimate hop (the last router before the destination machine): 41 of them appear to be located in Frankfurt, Germany because their names contain either the name of that city or airport/city abbreviations such as FRA and FFM. These 41 hops account for 3566 of the server IP addresses. One hop name, accounting for 743 IP addresses, contained the string *VIE*, which is the abbreviation of the Vienna airport in Austria. For the remaining penultimate hops, we had 5 IP addresses with no DNS name that account for 161 RapidShare IP addresses. Using IP-geolocation we pinned those addresses in a town in Germany close to Stuttgart (Leinfelden-Echterdingen).

In summary, our geolocation analysis suggests that RapidShare deploys all its server infrastructure at a single location that is in (or

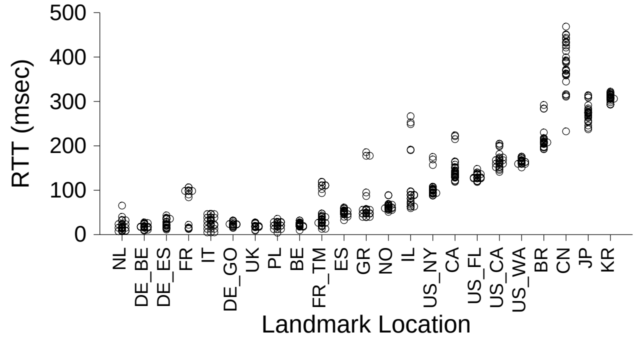



**Figure 11: Traceroute minimum RTTs (one for each RapidShare subnet) as measured from different Planetlab landmark hosts.**

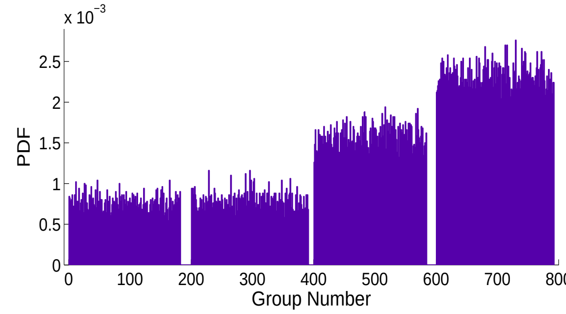



**Figure 12: Histogram of server group-IDs assigned to new upload requests.**

close to) Frankfurt, Germany. We will discuss the benefits of such a centralized infrastructure later in this section.

## 6.4 Content replication and server groups

Next, we estimate how many RapidShare servers host each file. To do so, we used the Tor anonymity network [9] as a geographically distributed network of clients. Specifically, we first collected almost 22,000 RapidShare URLs from public indexing web sites, and then we repeatedly requested those URLs for download using 421 different Tor exit nodes around the world (thus, each Tor node appeared as a different client to the RapidShare servers).

We observed two RapidShare servers in each download request. The first is used as the "indexing server" and it returns the server name that should be used for the download. The second is the actual "download server" that sends the requested file. Interestingly, the indexing server is always the same for a given file, while each file can be served by 12 RapidShare download servers. We refer to the dozen of servers that host the same file as a "server group". Further, it appears that all servers of the same group have two properties. First, the last byte of their IP address is the same. Second, those 12 servers belong to different subnets (of the same or different ISPs). As will be discussed later, each server group has a unique "group-ID" number that also appears in the server's name.

In summary, it appears that for each RapidShare hosted file, a unique indexing server redirects each client request to one of 12 download servers. To increase availability in the presence of ISP failures, and potentially to decrease transit fees, the download servers of the same group belong to different subnets (and often, to different ISPs).

## 6.5 Server Naming

RapidShare uses an interesting server naming scheme that allows easy identification of the upstream ISP and of the server group.

All server names start with the `rs` string. Then, the last byte from the server's IP address follows, either after reducing it by one (to have a starting point at zero) or after subtracting 1 and then adding 200, 400, or 600. The resulting number is the group-ID that the server belongs to. The next part of the name is the initials of the upstream ISP for that server. If there are several servers in the group that are connected to the same ISP, there is an additional number after the ISP initials.

For example, if a server has IP address *82.129.36.100* and its content provider is Cogent, its DNS name will be one of the following: *rs99cg.rapidshare.com*, *rs299cg.rapidshare.com*, *rs499cg.rapidshare.com*, *rs699cg.rapidshare.com*. The server's group-ID, in this example, will be 99, 299, 499, or 699. Thus, even though the last byte of the IP address can have only 256 values, the group-ID can take a much wider range of values. If there are two servers in that group connected to Cogent, the string "cg" will be followed by the number 1 or 2.

## 6.6 Server load balancing

In this section we explore the following two issues:
(1) Which server group will host a newly uploaded file?
(2) Which download server of that group will be used upon a download request?
To answer these questions, we performed a number of active measurements. When a user attempts to upload a file, the RapidShare service first responds with a list of 12 possible server names, part of the same server group. At that point, the group-ID is already determined and it can be inferred by the servers' name. The user can then select one of those servers and perform the upload. Note that RapidShare has no information about the size or type of the uploaded file when it determines the server group.

In our first experiment we performed 50,000 back-to-back upload requests to RapidShare (without performing the actual uploads). For each upload request we logged the returned group-ID. Figure 12 plots a histogram of the returned server groups. Observe that there exist four different ranges of group-IDs with different frequencies: 0-200, 200-400, 400-600 and 600-800. The first two have similar frequencies. It appears that the last two ranges correspond to large service expansions that RapidShare performed in Sep'08 (400-600) and in Mar'09 (600-800). Thus, more recently deployed servers have larger group-IDs, as one would probably expect. It is interesting that servers with larger group-IDs get a higher likelihood of upload assignments. Our interpretation of these results is that more upload requests are assigned to recently deployed servers, which should also have more available capacity, attempting to gradually balance the hosting load.

Our second experiment focuses on the download server selection process. Since each file is hosted by a group of 12 servers, how is the server that will handle a new download request selected? We performed one thousand back-to-back download requests for the same file and logged the returned download server. Figure 13 shows the histogram of the download servers for the corresponding group (group-ID: 717). Note that 10 out of the 12 servers have a very similar likelihood of serving the download requests. The two remaining servers are selected with lower probability. Server *rs717l3* is also used as the indexing server for this group, and it probably receives a lower download load due to its double role.

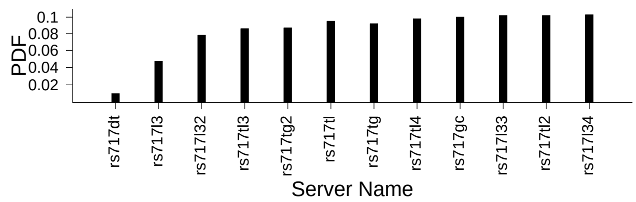



**Figure 13: Histogram of servers from the same group assigned to new download request.**

Server *rs717dt* has an even lower download load, but we do not understand why.

In summary, it appears that RapidShare uses simple load balancing rules, allocating more upload requests to more recently deployed servers, and assigning more download requests to servers that do *not* also act as the indexing server of the group. On the other hand, it seems that RapidShare does not try to control which ISP will be used for each download request (if they were doing so, the distribution in Figure 13 among the 10 download servers would not be uniform).

## 6.7 Discussion and comparison with CDNs

In this section, we explored a typical OCH service architecture Unfortunately, the service does not disclose such information and so we cannot determine the accuracy or correctness of our conclusions. Our results suggest that the architecture uses few thousands of servers (and the number keeps increasing) that are multihomed to several ISPs (some of them tier-1 transit providers). All servers are in the same location, probably close to Frankfurt in Germany. The servers are partitioned in groups of 12 servers each, while the servers of the same group belong to different subnets (and are often connected to different ISPs). Each file is hosted by one server group. It appears that upload requests are assigned to groups in a manner that considers the "deployment age" of each group. The actual upload server within the returned group, and the corresponding ISP, can be selected by the user. In terms of download requests, a user is directed to a specific download server from an indexing server, which is a member of the group that hosts that file. The assignment of downloads appears to be uniform across the download servers of the group.

The OCH architecture differs significantly from traditional Web Content Distribution Networks (CDNs), such as Akamai [1]. CDNs maintain mirror servers in many geographic locations, as they attempt to minimize the RTT between the user and the mirror server that will serve that user. It is not uncommon for a large CDN to be present at hundreds of geographical locations around the world. This difference with OCH is reasonable: a traditional CDN aims to minimize Web transaction delays and it is optimized for short TCP flows. An OCH service such as RapidShare, on the other hand, focuses on very large transfers that are less sensitive to delay. The centralized architecture of OCH is probably less expensive, easier to maintain, and it makes the migration of files between servers faster and less costly. Another difference is that the main value that CDNs offer to their customers is improved Web performance. For OCH services, performance is less important compared to content availability and the ability to share files inexpensively. Finally, the business models of CDNs and OCH services are very different. The former get their revenues from large content producers, while the latter get their revenues from individuals that subscribe to Premium accounts.

# 7. COMPARING RAPIDSHARE AND BITTORRENT

File sharing has traditionally used the p2p paradigm. p2p file sharing, mostly using BitTorrent, is the dominant source of traffic in the Internet today. In this section we explore whether the dominance of the p2p paradigm for file sharing can be challenged by the emerging OCH paradigm. Do OCH services provide significant benefits, in terms of performance or content availability, over p2p applications for file sharing?

To answer this question, we compare RapidShare, the leading OCH service, with BitTorrent, the leading p2p system, along the dimensions of download throughput and content availability. We think that these two factors are the most important from the typical user's perspective. Cost is another important factor, of course, when a user considers subscribing for a Premium RapidShare account. It should be noted that the following comparisons can only consider the current deployment of these two services; obviously, we cannot know whether the performance of RapidShare would deteriorate if that service was handling the same number of users as BitTorrent.

## 7.1 Download throughput

To compare performance, we manually downloaded 38 files from both services. The list of downloaded objects was randomly constructed from a number of indexing web sites, and it included a variety of files.The sizes of the selected objects range from 1.6MB to 2.85GB. Each content object was present in both services with approximately the same file size and quality. [7] We made sure that none of our downloads were illegal, requesting files that are not copyrighted. In the case of BitTorrent, we selected those torrents that had the larger population of seeders, to get the best download throughput. All downloads were done from the same client at FORTH. The client was connected to the public Internet through a 1Gbps access link.

For RapidShare, we used three different types of users. The first role is a **Premium user**. The second role is a **Free user**. Free users can download only one file at a time, and they are throttled to a throughput between 0.2Mbps and 2Mbps. Further, Free users have to wait for about 15 minutes between two successive downloads. In our measurements, we included these long wait times in the total download latency that a Free user experiences when she downloads an object that consists of several files. The third role is again a Free user that is able to change her IP address, through a new DHCP request, so that she can avoid the mandatory waiting period between consecutive downloads. We refer to such users as **Free-cheating**. [8]

Figure 14 shows the distribution of average throughput for each of the three user roles. Note that **Premium** RapidShare users enjoy a very high throughput compared to BitTorrent and the two other user roles. Indeed, more than 50% of the Premium RapidShare downloads got a throughput of more than 8Mbps - an order of magnitude higher than BitTorrent downloads. Of course, RapidShare Premium downloads are a paying service and many users would prefer to not pay for file sharing, given that p2p services are also free. The inner plot of Figure 14 suggests that the median download session experiences higher throughput with the Free-cheating RapidShare service than with BitTorrent. Indeed, 50% of the Free-

[7] In some cases, we could find the exact same object uploaded in both BitTorrent and RapidShare.

[8] It appears that **Free-cheating** is a popular behavior among OCH users. Many content indexing sites offer instructions on how to avoid the long waiting time that Free users have to experience. Of course, it is possible to do so when the user can acquire different IP addresses quickly and easily.

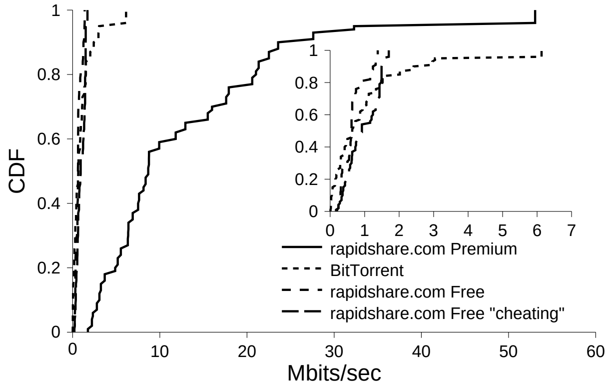



**Figure 14: Comparison of download throughput between three RapidShare user roles and BitTorrent.**

cheating RapidShare downloads receive more than 920kbps, while the Free RapidShare and BitTorrent median download throughput is about 600kbps. A small percentage of BitTorrent downloads however, receive much higher rates than Free-cheating RapidShare downloads. The fastest 10% of BitTorrent downloads receive more than 2.4Mbps.

## 7.2 Content availability

We next focus on the issue of content availability in the two services. To do so, we collected well-known lists of movies and songs, such as the Internet Movie Database [9], and searched for them in both RapidShare and BitTorrent. To verify that a specific object exists on RapidShare we first searched for it using Google, including the term "rapidshare" as a keyword together with the title of the object, and then examined the most relevant RapidShare link. We rejected search results that would not contain the full name of the object we search for. Then, we requested each collected URL from RapidShare to examine if it is still available for download (but without downloading the file).

To see whether a file exists on BitTorrent, we searched for it on `piratebay.org`, the most popular torrent search site today that hosts more than 600,000 torrents. The file was considered available if there was at least one seeder for it. [10]

Table 3 presents the related results. The second column shows the number of files contained in the object list. The third column is the number and percentage of objects that we were able to find on RapidShare using Google searches. The fourth column presents the number and percentage of objects found on BitTorrent. We see that in all cases, RapidShare has higher content availability than BitTorrent. For instance, in the "All Time USA Box Office" hits, 98.5% of the objects were found in RapidShare and 96.5% were found in BitTorrent. The difference between RapidShare and BitTorrent becomes more important for the less popular content. In the "Bottom 100 movies" list, RapidShare hosts 90% of those movies, while BitTorrent hosts only 50% of them. Users that are interested in this type of content seam to prefer RapidShare instead of BitTorrent. We speculate that users may prefer the simplicity of sharing files through a web page with URLs to those files in contrast to using a tracker and a specific file-sharing application.

[9] www.imdb.com

[10] This means that 100% of the file was available by at least one peer.

## 7.3 Summary

Our results suggest that RapidShare users enjoy a better file sharing experience than BitTorrent users. Indeed, Premium RapidShare users enjoy an order of magnitude higher download throughput than BitTorrent users. Even free RapidShare users have the benefit of more available content on RapidShare than on BitTorrent. Considering the actions a user has to follow to download an object, we believe that the process is equally simple in both services. In the case of BitTorrent, the user has to first search for the torrent file and then instrument her BitTorrent client to download that object. Similarly, a RapidShare user needs to first search for the object and download it through her browser. When an object spans several files, and thus several URLs need to be downloaded, there are simple "download managers" for Web browsers that can reduce the entire procedure to just a handful of mouse-clicks. Uploading files is much easier in RapidShare because the user only uploads a file in whole or in pieces. In BitTorrent one has to first find (or deploy) a hosting tracker, create the torrent file, and maintain an always-on host that will be serving the file until the swarm gets big enough. Considering these differences, file sharing user communities may have good reasons to use RapidShare instead of BitTorrent.

# 8. CONTENT INDEXING SITES

As previously mentioned, OCH services do not offer search or indexing capabilities. To fill this gap, a large number of content indexing Web sites offer such capabilities to OCH users. These indexing sites form several communities of users, ranging from general interest to very specific interest content. In this section we explore some of these sites to better understand how users search for objects in RapidShare, to examine the population of users that post download URLs on indexing sites, and to characterize the publicly visible RapidShare content in terms of type and copyright constraints.

## 8.1 Content uploaders

First, let us examine the community of RapidShare users that post download links on OCH indexing sites. To do so, we crawled four such sites shown in Table 4. The second column presents the number of OCH URLs provided by each indexing site. For example, `rapidshareindex.org` provides URLs to 54,327 objects hosted by OCH services; 36,522 of those are hosted by RapidShare. It is interesting that these 54,327 objects were uploaded by only 18 users! A very small number of uploaders is what we also observe in other indexing sites.

Figure 15 shows the percentage of posts by each user of three indexing sites. Again, we observe that a small number of users post almost all download URLs. For instance, in `rslinks.org`, only 5 users have posted more than 90% of the URLs, while in `rapidshareindex.com` a single user has posted (almost) all URLs.

As described in Section 2, files hosted by RapidShare become stale/invalid when the uploader requests to remove them, or when the uploader is a Free user and the file has either been downloaded 10 times or it has stayed inactive for a period of 90 days. Table 4 also shows the number and percentage of stale objects. We retrieved the state of each object by requesting it for download from RapidShare, and then checking for a valid download response. Stale files count for at most 19.3% of all files, and in one case for as little as 0.5%. This is in sharp contrast with file availability in p2p systems. For example, BitTorrent files have an average lifespan of only 9 days [13].

An interesting possibility is that the percentage of stale files is more community-specific, rather than service-specific. RapidShare

| List Name | Number of objects | Found on RapidShare | Found on piratebay.org |
|---|---|---|---|
| All Time Non-USA Box Office | 385 | 377 (97.9%) | 373 (96.8%) |
| Bottom 100 movies as voted by users | 100 | 90 (90%) | 53 (53%) |
| All Time USA Box Office | 408 | 402 (98.5%) | 394 (96.5%) |
| Top 250 movies as voted by users | 250 | 245 (98%) | 240 (96%) |
| All Time Worldwide Box Office | 346 | 338 (97.7%) | 336 (97.1%) |
| Top United States DVD Rentals for week ending 16 November 2008 | 50 | 50 (100%) | 50 (100%) |
| Amazon Best German films of all time | 25 | 21 (84%) | 19 (76%) |

**Table 3: Availability of movie objects in RapidShare and piratebay.org (BitTorrent).**

| Name | # of Indexed Objects | RapidShare Hosted Objects | # of Stale Files | # of Uploaders |
|---|---|---|---|---|
| `egydown.com` | 972 | 787 | 134 (17%) | N/A |
| `rapidmega.info` | 942 | 893 | 116 (13%) | 9 |
| `rslinks.org` | 12124 | 11841 | 64 (0.5%) | 21 |
| `rapidshareindex.com` | 54327 | 36522 | 7052 (19.3%) | 18 |

**Table 4: Description of four OCH content indexing sites.**

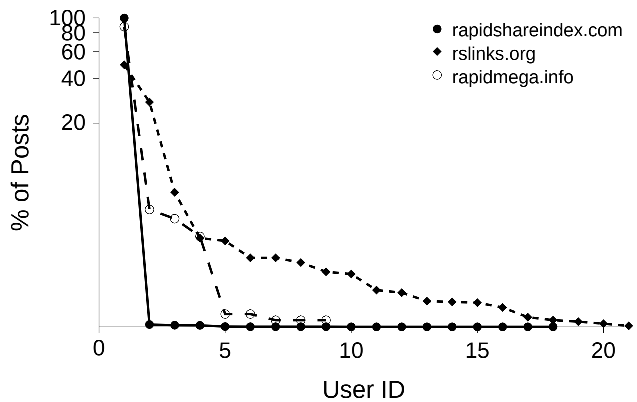



**Figure 15: Percentage of posts per uploader at three content indexing sites.**

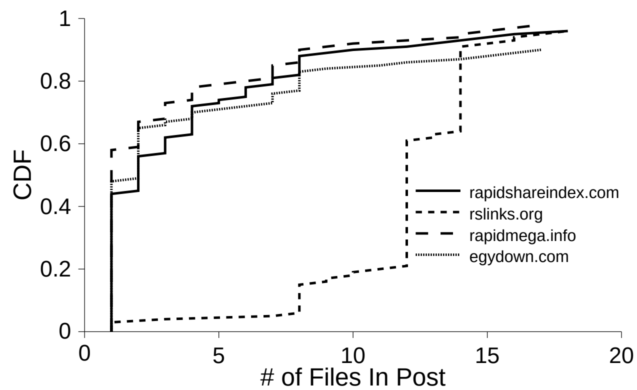



**Figure 16: Percentage of download URLs per indexed object.**

could also have a large fraction of stale files, especially for files uploaded by Free users. However, if a user community cares about a file, they would not let it become stale by successively uploading “fresh" versions. Based on the indexing sites we examined, we can say that this seems to be the case for RapidShare user communities.

## 8.2 Characterization of publicly visible content

Currently, RapidShare allows Free users to download files up to 200 MB. Thus, most of the large objects that are publicly shared are partitioned in files of that size (or smaller), and the user has to download a number of URLs. For instance, a full DVD of 2.4GB would be available through 12 URLs. In order to identify the type of objects that are available in RapidShare we crawled the previous four indexing sites and counted the number of download URLs per object.

Figure 16 shows the distribution of the number of URLs per object for each indexing site. At most 60% of the objects consist of a single URL. In the case of `rslinks.org`, this percentage drops to as low as 3%. A closer look to that site reveals that it is mainly used to share large video files (movies and TV series count for 80% of its total objects) and application/game CDs ( 20%).

To get a more clear picture of the type of objects that are publicly shared through OCH services, we manually examined the latest 100 objects seen at each of the previous four indexing sites. The results are summarized in Table 5. We see that video files, applications and audio collections are the largest categories - as expected. Again, `rslinks.org` is only used for movies.

## 8.3 Copyrighted content

Internet file sharing, through p2p networks or through OCH services, is constantly accused of copyright infringement. To get a rough estimate of the fraction of copyrighted content indexed in public OCH indexing sites, we manually examined the 100 most recent objects listed on each content indexing Web site. We classify an object as copyrighted if is a commercial movie, song, book or application, based on the official Web page of the corresponding content. Figure 17 shows the percentage of objects which appear to be copyrighted content. More than 84% of the 100 most recent objects, and in some sites 100% of the objects, appear to be copyrighted.

We note that OCH services, including RapidShare, are *not* responsible for any illegal file sharing and copyright infringements that take place using their infrastructure. RapidShare, in particular,

| Name | Games | Video | Apps | Books | Images | Audio |
|---|---|---|---|---|---|---|
| egydown.com | 11 | 19 | 65 | 1 | 4 | 0 |
| rapidmega.info | 0 | 45 | 1 | 1 | 0 | 53 |
| rslinks.org | 0 | 100 | 0 | 0 | 0 | 0 |
| rapidshareindex.com | 0 | 21 | 74 | 0 | 3 | 2 |

**Table 5: Classification of 100 objects from each content indexing Web site.**

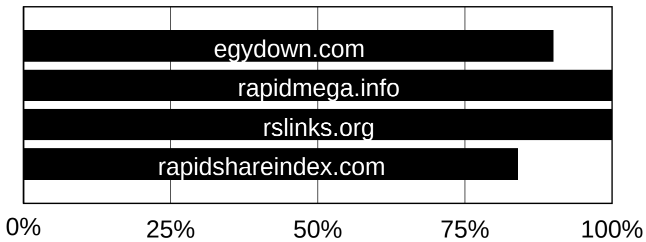



**Figure 17: Percentage of copyrighted material in a sample of 100 objects from each indexing site.**

aggress to host a file only if the uploading user declares that the content of that file is not copyrighted. [11] Further, RapidShare does *not* provide indexing and search capabilities for the content they store. Consequently, any copyright infringements that take place using their infrastruvture is solely a responsibility of the users that upload and download such content.

## 9. RELATED WORK

File sharing has attracted significant interest over the last ten years through the emergence of p2p protocols. The effect of p2p networks has been studied extensively [7, 13, 16, 23, 24]. Saroiu *et al.* studied the characteristics of participating peers in Napster and Gnutella, and observed significant heterogeneity and lack of cooperation among them [25]. Sen and Wang analyzed the characteristics and bandwidth requirements of p2p traffic in large networks [27]. Leibowitz *et al.* examined file popularity in the Kazaa network, and its effect on the total traffic volume [22].

The work of Gummadi *et al.* examined Kazaa's and Gnutella's file sharing workloads, and revealed significant popularity deviations between objects shared in Kazaa and Web objects mostly due to the "fetch-at-most-once" behavior of Kazaa users [12]. Karagiannis *et al.* on the other hand, showed that ISPs can benefit, in terms of cost, from locality-aware p2p distribution [18]. Our results for OCH services suggest that caching will not benefit ISPs since 75% of the files in our 4-5 month traces were downloaded only once.

Several recent studies focused on emerging Web 2.0 applications that host user-generated content. Gill *et al.* studied the characteristics of YouTube users' content in an academic network [10]. Cha *et al.* also studied user-generated content sites by crawling YouTube and Daum [4]. Focusing on video popularity, both studies found that the YouTube video workload is similar to traditional Web workload, thus end-users and organizations can benefit in throughput and reduced bandwidth demands, respectively, by employing caching policies or distribution through p2p systems. In late 2008 Cho *et al.* observed a slow but steady movement of residential users towards rich-media content, like YouTube, by comparing data from 2005 and 2008 [5]. During the time period of our study we did not observe a significant increase in OCH traffic, but we expect users to gradually turn to such services because they often offer better content availability and higher throughput than p2p networks.

Work on CDNs, like Akamai [1], has focused on the evaluation and improvement of the redirection and server selection algorithms, which result in lower latency and better quality-of-experience for end-users. Krishnamurthy *et al.* showed, in 2000, that the use of multiple distributed servers and caching in content distribution can improve end-to-end Web performance [20]. Johnson *et al.* measured the performance of two content distribution networks, Akamai and Digital Island, built over DNS redirection [17]. They note that both systems provide good but not optimal performance and conclude that their goal is not to select the best server, but to *not* select a bad server. Krishnamurthy *et al.* proposed a methodology to study the client-experienced performance using CDNs, and also performed an extensive study of the use of such systems [21]. The work presented by Saroiu *et al.* pursued a trace-based characterization of four Content Delivery Systems, Gnutella, Kazaa, WWW and Akamai [26]. Their results showed that Akamai consumed only a minimal amount of traffic (0.2%), while Gnutella, WWW and Kazaa followed with 6.04%, 14.3% and 36.9%, respectively.

## 10. CONCLUSIONS

Using a combination of passive monitoring and active probing we provide the first, to our knowledge, detailed study of OCH services, and of RapidShare in particular. Our results show that OCH services have reached a level of popularity at which *they produce a significant share of the daily Web traffic: a share that in some cases exceeds the traffic generated by YouTube and similar on-line video services*.

Most users download more than one file during their daily sessions. Interestingly enough, even over several months of observation, we saw very little locality of reference in the objects downloaded through OCH services, probably reflecting the diverse interests of file sharing users. These locality patterns imply that there will be no significant benefits by caching content near to users, which indicates that a centralized architecture would be appropriate for OCH services. Indeed, our experiments revealed that RapidShare employs a centralized, heavily multi-homed server infrastructure that is located at a single geographical location. RapidShare servers are grouped together into groups of 12, with each server in the group hosting the same set of files. Each server of the same group belongs to a different IP subnet and it is often connected to a different ISP. With frequent infrastructure expansions, RapidShare tries to keep its storage capacity evenly distributed among all server groups, by favoring recently added servers for new file uploads.

Comparing RapidShare with BitTorrent, we found that the former provides better performance than the popular p2p network. Our experiments suggest that *although Free RapidShare users experience similar throughput with BitTorrent users, Premium RapidShare users experience an order of magnitude higher throughput*. Furthermore, the amount of content shared over RapidShare is larger than the content shared with BitTorrent. Furthermore, the availabil-

[11] As is clearly stated in the "Terms of Use" agreement `http://rapidshare.com/agb.html`.

ity of content in RapidShare appears to persist for longer time periods than in BitTorrent. With a significant fraction of users willing to pay a small fee to enjoy premium service, OCH services appear likely to compete with BitTorrent as the leading file-sharing platform.

We also examined OCH content indexing sites, which are an essential component for file sharing using OCH services. We found that in OCH services, much like in p2p file sharing systems, *a very small number of users upload most files, which are often copyrighted content*, favoring audio albums, video movies, and applications.


## Acknowledgment

Demetris Antoniades and Evangelos P. Markatos are also with the University of Crete. Some of these experiments were done in collaboration with the SEEREN2 network. We deeply thank them for their support.


## 11. REFERENCES


[1] Akamai. `http://www.akamai.com`.

[2] Akamai Accelerated Network Program FAQ. `http://www.uvm.edu/~sjc/vbns/akamai/akamai-aanp-faq.pdf`.

[3] S. Bellovin. A Technique For Counting NATted Hosts. In *IMW '02: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurment*, 2002.

[4] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. I Tube, You Tube, Everybody Tubes: Analyzing The World's Largest User Generated Content Video System. In *IMC '07: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, 2007.

[5] K. Cho, K. Fukuda, H. Esaki, and K. Akira. Observing Slow Crustal Movement In Residential User Traffic. In *CoNEXT '08: Proceeding of the 4th Annual Conference on Emerging Network Experiments and Technologies*, 2008.

[6] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, and M. Bowman. PlanetLab: An Overlay Testbed For Broad-Coverage Services. *ACM SIGCOMM Computer Communications Review*, 2003.

[7] B. Cohen. Incentives Build Robustness in BitTorrent. In *Proceedings of the 1st Workshop on the Economics of Peer-2-Peer Systems*, 2003.

[8] A. Dhamdhere and C. Dovrolis. Ten Years In The Evolution Of The Internet Ecosystem. In *IMC '08: Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement*, 2008.

[9] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The Second-Generation Onion Router. In *SECURITY '04: Proceedings of the 13th conference on USENIX Security Symposium*, 2004.

[10] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. YouTube Traffic Characterization: A View From The Edge. In *IMC '07: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, 2007.

[11] D. Goldenberg, L. Qiuy, H. Xie, Y. Yang, and Y. Zhang. Optimizing Cost And Performance For Multihoming. *ACM SIGCOMM Computer Communication Review*, 2004.

[12] K. Gummadi, R. Dunn, S. Saroiu, S. Gribble, H. Levy, and J. Zahorjan. Measurement, Modeling, And Analysis Of A Peer-To-Peer File-Sharing Workload. In *SOSP '03: Proceedings of the 19th ACM Symposium on Operating Systems Principles*, 2003.

[13] L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, and X. Zhang. Measurements, Analysis, and Modeling of BitTorrent-Like Systems. In *In IMC '05: Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement*, 2005.

[14] ipoque. Internet Study 2007. `http://www.ipoque.com/news_&_events/internet_studies/internet_study_2007`.

[15] ipoque. Internet Study 2008/2009. `http://www.ipoque.com/resources/internet-studies/internet-study-2008_2009`.

[16] M. Izal, G. Urvoy-Keller, E. Biersack, P. Felber, A. Hamra, and L. Garces-Erice. Dissecting BitTorrent: Five Months In A Torrent's Lifetime. In *PAM '04: Proceedings of the 5th Passive And Active Measurement Conference*, 2004.

[17] K. Johnson, J. Carr, M. Day, and M. Kaashoek. The Measured Performance Of Content Distribution Networks. *Journal of Computer Communications*, 24(2):202–206, 2001.

[18] T. Karagiannis, P. Rodriguez, and K. Papagiannaki. Should Internet Service Providers Fear Peer-Assisted Content Distribution? In *IMC '05: Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement*, 2005.

[19] E. Katz-Bassett, J. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe. Towards IP Geolocation Using Delay And Topology Measurements. In *IMC '06: Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement*, 2006.

[20] B. Krishnamurthy and C. Wills. Analyzing Factors That Influence End-To-End Web Performance. *Journal of Computer Networks*, 33(1-6):17–32, 2000.

[21] B. Krishnamurthy, C. Wills, and Y. Zhang. On The Use And Performance Of Content Distribution Networks. In *IMW '01: Proceedings of the 1st ACM SIGCOMM Workshop On Internet Measurement*, 2001.

[22] N. Leibowitz, M. Ripeanu, and A. Wierzbicki. Deconstructing The Kazaa Network. In *WIAPP '03: Proceedings of the 3rd IEEE Workshop On Internet Applications*, 2003.

[23] E. Markatos. Tracing A Large-Scale Peer-To-Peer System: An Hour In The Life Of Gnutella. In *Proceedings of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid*, 2002.

[24] D. Nogueira, L. Rocha, J. Santos, P. Araujo, V. Almeida, and W. Meira Jr. A Methodology For Workload Characterization Of File-Sharing Peer-To-Peer Networks. In *WWC '02: Proceedings of the 5th IEEE International Workshop on Workload Characterization*, 2002.

[25] S. Saroiu, K. Gummadi, and S. Gribble. Measuring And Analyzing The Characteristics Of Napster And Gnutella Hosts. *Journal on Multimedia Systems*, 9(2):170–184, 2003.

[26] S. Saroiu, K. P. Gummadi, R. J. Dunn, S. D. Gribble, and H. M. Levy. An Analysis Of Internet Content Delivery Systems. In *OSDI '02: Proceedings of the 5th USENIX Symposium on Operating Systems Design And Implementation*, 2002.

[27] S. Sen and J. Wang. Analyzing Peer-To-Peer Traffic Across Large Networks. In *IMW '02: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurment*, 2002.

[28] N. Spring, L. Peterson, A. Bavier, and V. Pai. Using PlanetLab For Network Research: Myths, Realities, And Best Practices. *ACM SIGOPS Operating Systems Review*, 2006.