Paying for Piracy? An Analysis of One-Click Hosters' Controversial Reward Schemes

Tobias Lauinger[1], Engin Kirda[1], and Pietro Michiardi[2]

[1] Northeastern University, Boston, USA
[2] Eurécom, Sophia-Antipolis, France

Abstract. One-Click Hosters (OCHs) such as Rapidshare and now defunct Megaupload are popular services where users can upload and store large files. Uploaders can then share the files with friends or make them publicly available by publishing the download links in separate directories, so-called direct download or streaming sites. While OCHs have legitimate use cases, they are also frequently used to distribute pirated content. Many OCHs operate affiliate programmes to financially reward the uploaders of popular files. These affiliate programmes are controversial for allegedly financing piracy, and they were prominently cited in the criminal indictment that lead to the shutdown of Megaupload, once among the world's 100 largest web sites. In this paper, we provide insights into how much money uploaders of pirated content could earn on a range of direct download and streaming sites. While the potential earnings of a few uploaders are non-negligible, for most uploaders these amounts are so low that they cannot rationally explain profit-oriented behaviour.

Keywords: One-Click Hosting, Piracy, Uploader Income, Affiliate Programmes

1 Introduction

Piracy is the most common illicit activity on the Internet. Every day, millions of people use P2P networks or One-Click Hosters (OCHs) such as Hotfile, Rapidshare and formerly Megaupload to share copyrighted content without permission. File sharing based on OCH works in a division of labour: OCHs provide the storage but no search functionality, and external direct download or streaming sites host searchable repositories of download links pointing to the OCHs.

OCHs are large businesses financed through advertisement and subscription fees; several of them are among the 100 largest web sites worldwide. Because OCHs have various legitimate use cases, they claim immunity against their users' copyright infringements under the U.S. Digital Millennium Copyright Act.

However, many OCHs also operate controversial affiliate programmes in order to attract new paying members. These affiliate programmes financially reward uploaders based on the number of downloads and member subscriptions that they generate. For instance, Megaupload used to reward one million downloads with $ 1,500 and WUpload used to pay up to $ 40 per one thousand downloads. These

affiliate programmes are controversial for allegedly encouraging users to upload copyrighted content and thereby funding piracy. For instance, Megaupload's former affiliate programme and their knowledge that affiliates uploaded pirated content were a central element of the criminal indictment[3] that lead to the seizure of Megaupload's assets, the detention of its operators, and the shutdown of the site on 19 January 2012.

In this paper, we investigate how much money uploaders can earn by illegally uploading pirated content and posting download links on a range of direct download and streaming sites. The order of magnitude of an uploader's income tells us whether the affiliate programme and the associated rewards should be considered as a major factor in the uploader's motivation, or if they could be seen as just a minor concomitant effect.

Measuring uploader income is a challenging task: Almost no OCH reports how often a file was downloaded, and most direct download and streaming sites do not display how often a download link was clicked. Furthermore, even if these data are known, nothing reveals whether an uploader actually participates in an OCH's affiliate programme.

We tackle this problem in the following way: We crawl three large direct download/streaming sites that make click data available. Using the click data, we compute an uploader's maximum income for the links posted on the site under the assumption that every click generated a valid download, and that the uploader participated in the affiliate programme. In order to estimate how many clicks correspond to an actual download, we correlate the click data with the number of downloads on the few OCHs that make download data available.

Our results show that most uploaders earn next to nothing; they do not exhibit apparent profit-oriented behaviour. However, we also observe that a handful of uploaders upload large numbers of files each day and generate so much traffic that they could earn up to a few hundred dollars per day. For these uploaders, at least some degree of profit-oriented behaviour is probable.

Our findings have implications on proposed anti-piracy measures such as the U.S. draft bill SOPA and similar projects in other countries that aim at interrupting the revenue stream of piracy: Such measures, by definition, can affect only profit-oriented actors. Given that we observe a large number of altruistic uploaders, these measures run the risk of having only little effect overall.

In this paper, we make the following contributions:

- We are the first to use large-scale empirical data to estimate the distribution of uploader income through affiliate programmes. We contrast the income with indicators for the effort invested by uploaders. This tells us about the motivations of uploaders with respect to profit seeking or altruism.
- We are the first to provide insights into how the shutdown of Megaupload and the associated cancellations of other OCHs' affiliate programmes affected illegal uploader income. This gives us ground truth to judge the success of anti-piracy measures that aim to curb piracy by removing financial incentives.

[3] Superseding indictment, U.S. v. Kim Dotcom et al., 1:12-cr-00003-LO (E.D. Va., Feb. 16, 2012) at ¶ 58; ¶ 73 g–j, v, y, bb, jj, pp, qq, uu, ppp, qqq, www, xxx; and ¶ 102.

2 Background

One-Click Hosters (OCHs) have various legitimate use cases, such as storing backups or exchanging large files instead of sending them as email attachments. Because the purpose of this paper is to measure illegal uploader income relating to piracy, we focus the background information given in this section on illicit file sharing and on ways of monetising pirated content.

2.1 OCH-Based File Sharing and Streaming

One-Click Hosters such as Rapidshare, Megaupload, Hotfile or Mediafire provide web-based storage for potentially large files. Users can upload files through a simple web interface. For each uploaded file, the OCH provides a unique download link to the uploader. Because most OCHs do not make uploaded files public or offer search capabilities, uploaders seeking to publish their files need to post the corresponding download links on third-party web sites. There is a great variety of such sites, ranging from general-purpose discussion boards and blogs to more specialised content indexing sites, so-called direct download sites. These sites offer a catalogue of links, supplied by site staff and sometimes independent users, including categories such as movies, TV shows, games, music, ebooks, and software. So-called streaming sites index movies and TV shows using an embedded video player provided by OCHs such as Megavideo, VideoBB and Putlocker. In the following, we will use the term link or indexing site to refer to all types of "underground" web sites that are specialised in supplying links to pirated content hosted on OCHs.

As Fig. 1 shows, relationships between OCHs and indexing sites can be complex: Some uploaders spread their links over many indexing sites. An individual indexing site typically contains several copies of the same content hosted on different OCHs, and sometimes even several "mirror" copies of the same file hosted on the same OCH. Instead of posting the original download link, some uploaders use URL shorteners or "link protection services". The purpose of these services is to protect download links against automated extraction by web crawlers run by copyright holders to automatically take down files that infringe their copyright. Sometimes, these services are also used to better monetise links, such as by displaying advertisements before redirecting the user to the OCH.

2.2 OCH Affiliate Programmes

One-Click Hosters usually offer a free, advertisement-based service and a premium subscription service. In order to convert free users into paying members, the free service is artificially limited in the bandwidth, and free users need to wait between consecutive downloads. According to the indictment[4], Megaupload received at

[4] Superseding indictment, U.S. v. Kim Dotcom et al., 1:12-cr-00003-LO (E.D. Va., Feb. 16, 2012) at ¶ 4.

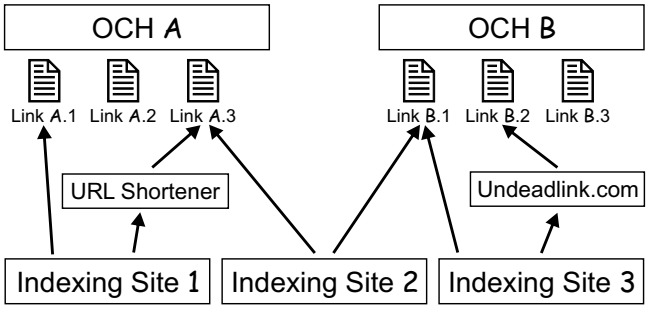


Fig. 1. The OCH ecosystem: Indexing sites can refer to a range of OCHs, the same download link can be posted on several indexing sites, mirror copies of the same file can be hosted on different OCHs or even on the same OCH, and links can be hidden behind a layer of redirection by using URL shorteners, for instance.

least 150 million dollars in subscription fees and 25 million dollars for advertising between September 2005 and 5 January 2012.

There are hundreds of competing OCHs. In order to attract user traffic and generate membership sales, most OCHs offer affiliate programmes for uploaders and indexing sites. Affiliate programmes differ widely in the amounts paid, but they are always a combination of these basic building blocks:

Pay Per Download (PPD). A small amount of money is paid for each (full) download, such as $ 15 for 1000 downloads. Often, the amount differs according to the country of the downloader; Table 1 shows as an illustration the rates that were paid by WUpload until late November 2011. Some OCHs use different affiliate "levels" to weigh the payouts according to the past performance (which includes the conversion rate: premium sales per traffic). In most cases, only uploaders can participate in PPD programmes.

Table 1. PPD rates for WUpload, per 1000 downloads, retrieved on 30 October 2011. Country group A: *US, UK, DE*. Group B: *AU, AT, BE, CN, DK, FI, FR, IE, IT, JP, NL, NZ, NO, SA, SG, SE, CH, AE*. Group C: *BR, BG, CY, CZ, GR, HU, IR, KW, LV, LT, LU, PL, PT, QA, RO, RU, ZA, ES, TR*. Group D: *All others*. WUpload discontinued the programme in late November 2011.

Size/Country	A	B	C	D
1−50 MB	$ 5	$ 3	$ 2	$ 1
51−100 MB	$ 12	$ 8	$ 5	$ 3
101−250 MB	$ 19	$ 15	$ 12	$ 5
251−400 MB	$ 27	$ 20	$ 18	$ 7
401−2048 MB	$ 33	$ 26	$ 22	$ 10
2048+ MB	$ 40	$ 28	$ 24	$ 12

Pay Per Sale (PPS). A commission is paid for each premium sale or extension of subscription ("rebill"). The amounts paid are the same across all countries, and both uploaders and website owners can participate. For instance, WUpload used to reward uploaders with 70 % of new premium subscriptions in their PPS-only affiliate programme. Webmasters could earn 10 % of the sales to visitors that came from the webmaster's site.

Sometimes, uploaders can choose from different "formulas" such as PPD only, PPS only, or 50 % of PPD + 50 % of PPS. Not surprisingly, new OCHs tend to pay more generously, either through higher rates, or by running "promotions" during which each affiliate's payout is doubled, for instance. In the aftermath of the Megaupload shutdown, many OCHs discontinued their affiliate programmes (including VideoBB, Fileserve and Filepost), converted their affiliate programme into PPS only (Uploaded), disabled file sharing functionality (Filesonic and later WUpload) or decided to shut down voluntarily (X7).

Indexing sites can generate income through advertising, the PPS component of OCH affiliate programmes, by uploading files themselves (and fully leveraging OCH affiliate programmes), and sometimes by collecting donations. For the purpose of this paper, however, the revenue of OCHs and indexing sites is considered out of scope as we focus on uploader income through the PPD component of OCH affiliate programmes.

3 Methodology

Estimating uploader income is a difficult task because the sale and download transactions rewarded in affiliate programmes cannot easily be observed from an outsider's perspective. Sales data are kept secret by all OCHs, and only a few OCHs report the number of downloads of each file. A few indexing sites display how often a file has been "downloaded", which in reality means how often the link has been clicked.

In this paper, we focus on uploader income through the PPD component of affiliate programmes because it is the only type of income that we can measure empirically and on a large scale. We estimate uploader income by extracting the links posted on three large indexing sites along with click-through counters that are displayed on these sites. Whenever possible, we compare this data with ground-truth download data that a few OCHs supply in their APIs.

3.1 Data Sources

The income through PPD depends on the number of files an uploader has, how often each file is downloaded, and what amount the OCH pays for each download. The latter information can be obtained from the OCHs' websites since most OCHs openly advertise their affiliate programmes, if they have one, and allow any user to join. Data about the number of downloads is much more difficult to obtain; most OCHs and indexing sites do not make it publicly available.

To prepare our study, we visited the most popular indexing sites in a range of countries and checked what metadata they published. For our study, we retained three sites that counted the number of clicks of each link:

- `Dpstream.net` is the largest streaming site in France and contains movies and TV shows. For our study, we crawled the movies section only. While the site did not make any click data available, around half of the movies were hosted on VideoBB, an OCH that reports view data for their videos. (Today, the site uses a different set of OCHs.) In our analysis, we use the ground-truth view data instead of the (unavailable) click data.
- `Iload.to` is the largest direct download site in Germany (it is preceded only by two streaming sites). It consists of a directory of links that are provided by staff, a separate exchange board with user uploads, and various other community functions. We focussed on the section with staff uploads because it displayed the number of clicks of each link. The content published on the site includes movies, TV shows, music, ebooks, games, software, and pornographic material.
- `Redlist-ultimate.be` is a Belgian file sharing community with a large index of movies, TV shows, music, ebooks, games, and software. Links can be submitted by registered users only, and there are various filter rules and staff intervention to keep the index organised. Each link is annotated with additional information such as the name of the uploader and the number of clicks. The site is not as popular as the two other sites, but it publishes valuable information about registered users, such as the number of uploads and downloads, and the total time spent logged in. Out of the registered users, 79 % report France as their country.

The vast majority of the content posted on these indexing sites is being commercially exploited and is sometimes even available before the official release date in stores. During our measurements, we witnessed only a dozen content items that seemed to be shared legitimately, and their popularity was low compared to the remaining (pirated) content on the sites.

3.2 Data Sets

To obtain data sets with the links posted on indexing sites, along with the corresponding click data, we performed a series of crawls on the three indexing sites mentioned above. Table 2 lists the key characteristics of these three sites and the data sets that we extracted from them.

For `dpstream` and `redlist`, we carried out a series of *full crawls* during which we extracted all the existing content and metadata. (Our `dpstream` data set is restricted to VideoBB links in the movies section of the site.) We repeatedly performed full crawls during one month. For `redlist`, we performed an additional series of crawls in March 2012, slightly less than two months after Megaupload had been shut down, to assess the impact of this event on the file sharing ecosystem.

Due to the very high number of content objects (movies, TV show episodes etc.) indexed on `iload`, a full crawl would have taken too long to complete. Instead,

Table 2. The indexing sites crawled for this study and the types of data available on these sites. Media content is broken down into downloads and streams. Click data is provided by the indexing site; OCH views (or downloads) are ground truth collected from the respective OCH. `Dpstream` is limited to movies hosted on VideoBB and uses OCH views instead of click data. For `iload`, clicks and payout refer to the first 30 days in the lifetime of all objects that are added on a single day.

Name	`dpstream`	`iload`	`redlist-oct`	`redlist-mar`
Alexa Rank	1507 *(FR: 70)*	2976 *(DE: 144)*	15405 *(FR: 735)*	
Downloads	✗	✓	✓	
Streams	✓	few	✗	
Crawl Start	18 Feb 2011	4 Apr 2011	3 Oct 2011	8 Mar 2012
Crawl End	23 Mar 2011	10 Jul 2011	5 Nov 2011	22 Mar 2012
Crawl Type	full	new content	full	
Click Data	✗	✓	✓	
Uploader Data	✓	few	✓	
OCH Views	✓	few	✗	few
# Content	10,950 total	421 added/day	114,475 total	43,418 total
# Links	11,026 total	7,674 added/day	358,297 total	109,492 total
# Clicks/day	16,349	223,691 (future)	140,996	148,090
$ Payout/day	32.70	1,010.50 (future)	184.79	1028.48
Comments	films/VideoBB	future 30 days	pre/post Mega* shutdown	

we crawled only the *new content* that was added to the site: We requested the site's RSS feed every hour to discover new content. At the same time, for all discovered objects, we periodically (and repeatedly) retrieved the associated pages to track the evolution of the number of clicks. We ran this experiment for around three months, until `iload` stopped publishing click data.

Our crawler was capable of detecting more than 500 different link types from 300 different OCHs. In order not to distort the click count when extracting links from the indexing sites, the crawler kept track of its requests and we adjusted the final click data accordingly. For each discovered link that referred to an OCH that made download data available, we furthermore retrieved the number of downloads from the OCH's API every two days.

To extract information about the OCHs' affiliate programmes, we visited the websites of more than 50 OCHs used on the three indexing sites in October 2011 and again in March 2012. Several OCHs modified their affiliate programmes during our study. For instance, Megaupload discontinued their affiliate programme in summer 2011, thus we use their rates for `iload` but not for `redlist`.

The amounts paid per download are often differentiated by the file size and by the country of the downloader, as illustrated for WUpload in Tab. 1. To look up a consistent payout value for all files, we make the following assumptions: For links found on `dpstream` and `redlist`, we assume all downloaders to be located in France; for `iload`, we use the payout amounts for Germany. These assumptions correspond to the countries where most of the sites' users come from.

We furthermore assume a constant file size of 101 MB because this is a common (and conservative) value on file sharing sites [9]. For streaming links, we assume a video length of 90 minutes because most of the streaming links found in our data sets correspond to movies. (Most TV shows only have download links.)

3.3 Ethics

All the data in our data sets was collected from public sources that are accessible to every Internet user. Our data sets contain no IP addresses or real names; the most private information that we possess are the (publicly visible) user names of the users who posted links, and in some cases the user names of the file uploaders. However, these names are freely chosen by the users and we have no means to map these user names to a real-world identity. Therefore, our analysis does not negatively affect the privacy of any individual uploader.

3.4 Metrics

The direct way to infer the income of uploaders is to use view or download data supplied by the OCH and multiply it with the PPD amount. Unfortunately, only the `dpstream` data set has a representative amount of OCH-provided view data. For the other data sets, we infer the income indirectly through the number of clicks observed on indexing sites.

To approximate an uploader's income generated by PPD programmes, we define the *value* v of a link $l \in L$ as follows:

$$v_\alpha(l) = \text{clicks}(l) \times \alpha \times \text{payout}(\text{och}(l)) \ , \tag{1}$$

where $\text{clicks}(l)$ is the amount of clicks reported on the indexing site for a given time frame, α is the click-download ratio, that is, the fraction of clicks that result in a valid download, and $\text{payout}(\text{och}(l))$ is the amount of money paid by the OCH of the link for one download. Note that the value of a link is different from the uploader's income because it refers to *potential* income that depends, among others, on the actual value of α. We discuss this issue in more detail in Sect. 3.5.

We express the number of clicks as daily averages. For `dpstream` and `redlist`, our data sets contain a sequence of full crawls, as shown in Fig. 2 for two crawls. In the regular case, we have one observation of c_d clicks in the first crawl at time t_d, and another observation of $c_e \geq c_d$ in the second crawl at time t_e. We compute the average number of clicks per day as $\frac{c_e - c_d}{t_e - t_d}$. Note that we consider only links present in the first crawl; links that are added at a later time will be discarded. Similarly, if a link is deleted before we can take a second snapshot, we cannot compute the number of clicks. On `redlist`, a full crawl took between six and ten days to complete.

The sites that we have crawled contain tens to hundreds of thousands of links, and not all of the links receive a click between two successive crawls. Therefore, we use the first crawl to determine the set of links that will be considered, and the

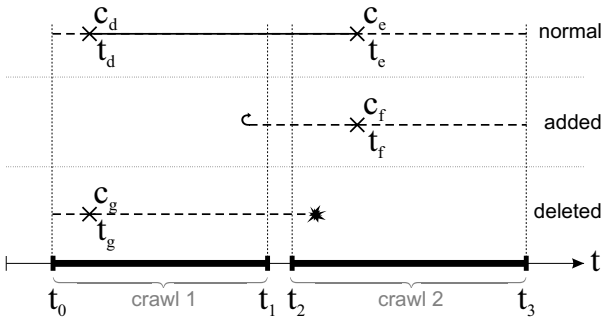


Fig. 2. Click computation for *full* crawls (`dpstream` and `redlist`).

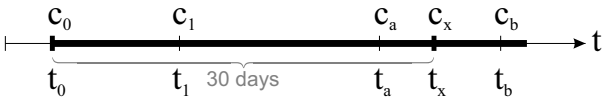


Fig. 3. Click computation for crawls for *new* content (`iload`).

click counts observed in the *last* crawl to compute daily averages[5]. This tradeoff permits us to improve accuracy for unpopular content while not penalising popular content (with quickly decaying popularity) too much.

While the data for `dpstream` and `redlist` covers the existing content on the site (both old and new), the data set for `iload` contains only new links that were added to the site. For this site, we use a different definition of "average daily clicks". As diagrammed in Fig. 3, we start tracking a new link when it is published at time t_0, and we take successive snapshots of the number of clicks c_i at time t_i, $i \geq 0$. Our goal is to estimate how many clicks c_x a new link generates in the first $t_x = 30$ days of its lifetime. In contrast to the full crawls, the click count snapshots are taken in different time intervals, according to the degree of utilisation of the crawler. In order to obtain an accurate estimate of c_x, we perform linear interpolation between the latest click count c_a observed before t_x, and the earliest click count c_b observed after t_x. The estimated value for the click count after thirty days is then $c_x = c_a + (c_b - c_a) \cdot \frac{t_x - t_a}{t_b - t_a}$.[6] This metric defines the value of a link with respect to the number of clicks that the link will generate in the first thirty days of its lifetime. We can use this metric to compute for each day how much *future value* an uploader generates by adding new links to the site. We can furthermore average over all days to obtain the *daily future value* generated by adding new links to the site.

To summarise these metrics, for `dpstream` and `redlist`, we compute for each existing link how many clicks it receives per day. On `iload`, we characterise the

[5] In the case that a link is deleted in the meantime, we use the latest click observation that we have, but divide by the total time span between the first and the last crawl, that is, around 23 days for `redlist-oct`.

[6] If the link is deleted before t_x and we have no observation c_b, we simply use $c_x = c_a$.

dynamics of the page by computing not only how many links are added to the site each day, but also how many clicks these new links generate in the first thirty days of their lifetime.

3.5 Limitations

Due to the methodology we have chosen, we can compute the distribution of uploader income, but we cannot know if an uploader actually participates in the affiliate programme. Yet, previous work has shown [4] how rapidly OCHs that discontinued their affiliate programmes lost user traffic, which suggests that those affiliate programmes were a driving factor behind these OCHs' popularity.

In most of our data sets, the click-download ratio α is unknown. If users post their links on various indexing sites in addition to the three sites of our data sets, it is possible that $\alpha > 1$. On the other hand, $\alpha < 1$ if visitors click on the link without downloading the file, as it can happen when the file was deleted from the OCH. Furthermore, some OCHs count only completed downloads and take into account only one download per day and IP address. We address this issue by using OCH-provided ground truth on download data in `dpstream`. For `iload` and `redlist`, we compute the *maximum value* of a link on the indexing site as the payout generated by the indexing site's traffic with $\alpha = 1$. This definition ignores the payout contribution due to traffic from other indexing sites and assumes that every click generates a download. In the few instances where both click and download data is available, we can estimate α and scale down the maximum link value to obtain a more realistic approximation.

Since content uploaders and link posters are not necessarily the same person, what we estimate in this paper is *how much the links are worth* that users post on indexing sites. We refer to this as (potential) uploader income because it is what uploaders can make if they are interested.

For practical reasons, we need to make a range of simplifying assumptions, such as a static file size and downloader country. Furthermore, we do not consider any payout threshold (which can be up to $ 200 for some OCHs) that prevents uploaders with low income from being paid. For this reason, the results that we provide in this paper should be seen as best-effort approximations that hold on the long term.

4 Results

Each of our data sets provides us with insights into different aspects of the monetisation of pirated content: `Dpstream` gives us a global view on the distribution of uploader income based on ground-truth data (Sect. 4.1). `Iload` shows the value of individual links and assesses their depreciation over time (Sect. 4.2). `Redlist` allows us to characterise individual uploaders, including the effort that they put into their activity and their importance to the functioning of the site (Sect. 4.3). The second `redlist` crawl furthermore provides us with insights into the effects of the Megaupload shutdown on the money-making opportunities in the file sharing ecosystem (Sect. 4.4).

4.1 Uploader Income

To compute the income distribution of uploaders, we consider the VideoBB links posted in the movies section of `dpstream`, France's largest piracy-based streaming site. VideoBB links are available for approximately half of the movies. Because we use view data retrieved directly from VideoBB, we expect our results to be very close to what VideoBB actually paid to participating affiliates.

Figure 4 ranks the site's users by their income and plots the users' share of the site-wide income and total number of VideoBB links. From a global point of view, the income is concentrated on a few uploaders. For instance, the top 4 uploaders earn more than 30 % of the total income. The top 50 users receive almost 80 % of the total income and provide around 70 % of the links. One could argue that anti-piracy measures targeting the top 50 uploaders seem promising as the site would lose a large portion of its links. While true in this specific case, we show in Sect. 4.3 that this intuition is wrong in a more general scenario.

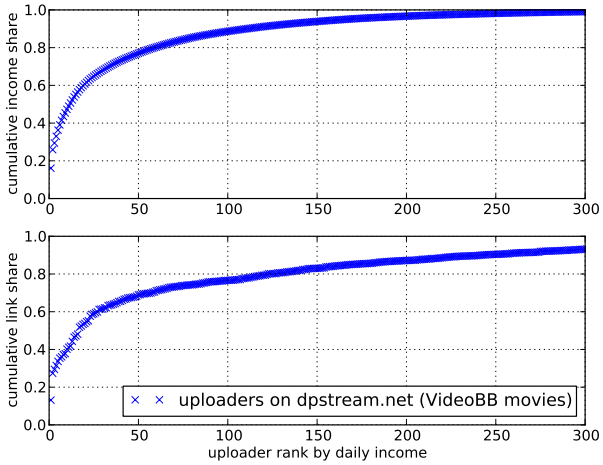


Fig. 4. Value and link share in `dpstream`: Associates each of the 585 uploaders, ranked by their income, with their share of the site-wide income of $ 32.70 per day (above) and the fraction of the total 11,026 links that they provided (below). A few users generate most of the value and most users earn almost nothing. While the users with the highest income also provide most links, the share of links provided is disproportionately small compared to the income share.

With a site-wide daily payout of $ 32.70, the potential earnings of individual uploaders are surprisingly low: 60 % of the users post content that is worth less than one cent per day, and even the top uploader can earn only $ 5.26 per day. While the low income in absolute terms appears to preclude profit-oriented uploader behaviour, the `dpstream` data set does not reveal much about the effort

associated with an uploader's activity, that is, how often an uploader needs to provide new links in order to have a steady stream of revenue. Furthermore, it is unclear where uploaders are based and whether the amount of their income should be assessed according to western standards or to those of a developing country. We come back to these issues in the following sections.

4.2 The Value of a Link

We use the `iload` data set to analyse the popularity of content objects and the choice of OCH made by the uploaders. The popularity of content objects such as movies, episodes of TV shows, games or ebooks is important because only popular objects can yield any significant payout. The choice of OCH is crucial because it determines how well an object's popularity can be monetised.

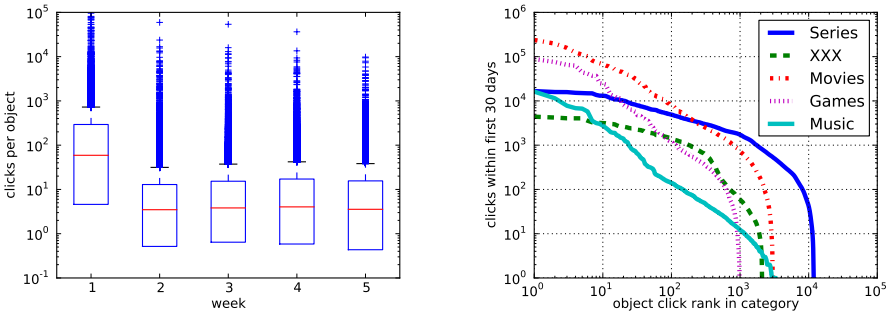
In the data set, we keep track of new objects being added to the site. Because `iload` is specialised in the timely publication of releases leaked by the Warez Scene[7], we can assume that the content is "fresh" when it is posted and analyse how its popularity evolves over time. Figure 5(a) shows a box plot of the weekly click distribution per content object. The popularity of content decreases very quickly: While the median number of clicks is 59 in the first week, it drops to less than 5 clicks per week in the following weeks. The 99^{th} percentile drops from 4,383 in the first week to 300−366 in the following weeks. Even though the click distribution exhibits outliers that continue to receive more than 10,000 clicks per week, the vast majority of content becomes "worthless" after only one week. As a result, uploaders wishing to make money need to regularly post fresh content.

When looking at the popularity of individual objects as shown in Fig. 5(b), it becomes clear that the site posts a lot of relatively unpopular content. For instance, 25 % of all new movies receive less than 100 clicks in the first 30 days; only 3 % of the movies receive more than 10,000 clicks in the same time span. An object with a few hundred clicks per month makes a couple of dollars at most and cannot generate any noticeable income through advertising either. Note furthermore that each object uploaded on `iload` corresponds to several alternative links and mirror copies on at least a half dozen OCHs. We argue that for such objects, even if automated, the cost of uploading can hardly be amortised by the income generated by these objects. The reason for posting unpopular objects might rather be a matter of prestige.

This issue becomes even more acute at the level of granularity of individual links: Within the first 30 days, a single link can make up to $ 335.29. However, only the top 20 links achieve a potential payout of more than $ 100 in their first 30 days. The median, even if considering only links that received at least one click, is merely 2 cents for 30 days. Only by adding 421 new objects (7,674 new links) every day can `iload` achieve a significant income: For all content posted at most 30 days ago, the combined PPD income is up to $ 1,010.50 per day.

Figure 6 breaks down the recent content objects' clicks and value by OCH. Although Megaupload, Uploaded, X7 and Fileserve are the most popular OCHs

[7] For an introduction to the Warez Scene, refer to [8], [7] and [3].



(a) Content hotness on `iload`: 75 % of the objects receive less than 20 clicks per week once the object is older than one week.

(b) Object popularity on `iload`: Most of the new objects receive only few clicks within the first 30 days.

Fig. 5. Content object popularity on `iload`: Popularity (a) per week and (b) by category.

with uploaders, only Megaupload is equally popular with downloaders. In fact, the OCHs most popular with downloaders pay the least competitive rates or nothing at all to uploaders. (OCHs without payout account for 24.1 % of the links and 30.6 % of the clicks.) Most links (95 %) seem to be uploaded by staff of the site and their policy to provide links to OCHs with low or no payout reduces the potential income of the site. While this finding might suggest that the site does not attempt to maximise PPD profit, one should keep in mind that the overall popularity of the site might suffer if the users' favourite OCHs are not offered, which is particularly important for users who have paid for premium services on one OCH.

So far, the value of links computed for `iload` was based on $\alpha = 1$. `Iload` posts a small number (16,553) of VideoBB streams. For these links, we estimate $\alpha \approx 0.40$ by linear regression as shown in Fig. 7(a) (correlation coefficient 0.69, $R^2 = 1$), which means that the actual payout is significantly lower than what the click count alone would suggest. In the `redlist-mar` data set, we estimate $\alpha \approx 0.73$ (correlation coefficient 0.65, $R^2 = 0.90$) for Files-Save (Fig. 7(b)) and $\alpha \approx 1.75$ (correlation coefficient 0.65, $R^2 = 0.99$) for Fufox (Fig. 7(c)). Here, $\alpha > 1$ suggests that those links are also posted on sites other than `redlist`. Note, however, that none of the latter two OCHs rewards uploaders with cash.

Many large downloads are split into smaller parts. In order to reassemble the original file, the downloader needs to download all parts of such a group of links. However, Fig. 7(d) shows that there is a difference of 32 % (correlation coefficient 0.87, $R^2 = 1$) between the link with the lowest and the link with the highest click count. In other words, only 32 % of the users who were interested in a file proceeded to download it entirely. These results illustrate that in most cases, $\alpha = 1$ induces a conservative upper bound on the actual number of downloads.

To summarise, most objects make money only for a limited time (one week) and need to be replaced regularly in order for the uploader to earn a regular

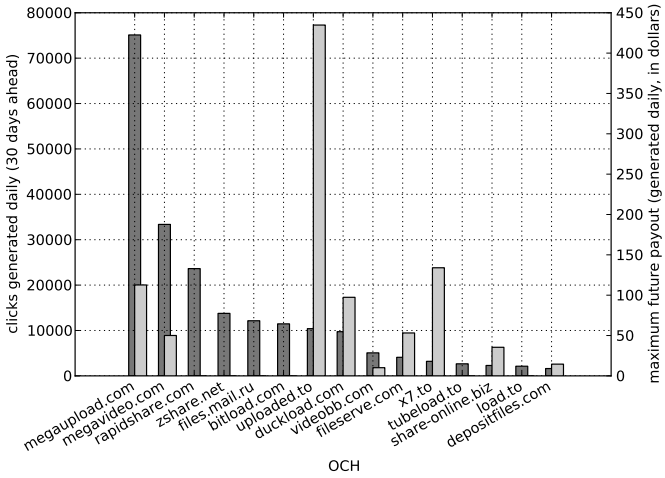


Fig. 6. Clicks (dark grey) and value (light grey) of links on `iload` for the top 15 OCHs: Some OCHs pay much higher rates than others, and the site also makes use of OCHs that do not pay any rewards at all.

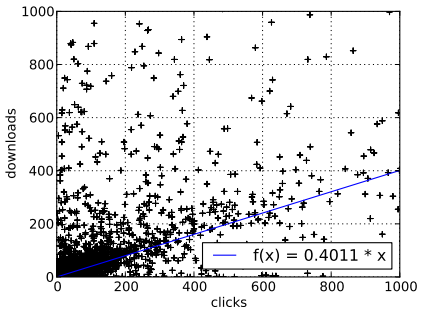
income. The choice of OCHs and especially the large quantities of highly unpopular objects suggest that `iload` is not maximising its profitability.

4.3 Characteristics of Top Content Uploaders

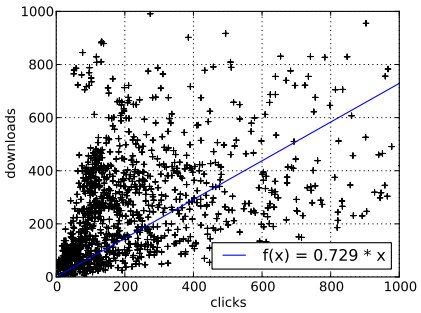
`Redlist` contains insightful data about the users registered on the site. We use this data to answer whether uploading can be profitable—how much work in terms of uploaded files and online time the best earning uploaders carry out, and how much their potential income might be worth to them, by looking at which countries these uploaders come from. Furthermore, we investigate how essential the top uploaders are to the functioning of the site.

We use the `redlist-mar` data set because it is more recent and reflects better the current state of the site after the shutdown of Megaupload. It contains 101,300 registered users, out of which 7,960 logged in at least once during the week of the crawl and 275 posted at least one link. The median number of links that downloaders click on is slightly larger than the median number of links that uploaders post (Fig. 8). However, the activity distribution of uploaders is more heavy-tailed with a few uploaders posting more than 100 links every day. Similarly, 30 % of all active uploaders and 70 % of the 50 highest earning uploaders spend more than one hour logged per day, whereas this is the case for only 4 % of the users who do not post links. These numbers illustrate that the top uploaders invest a significant effort into their activity.

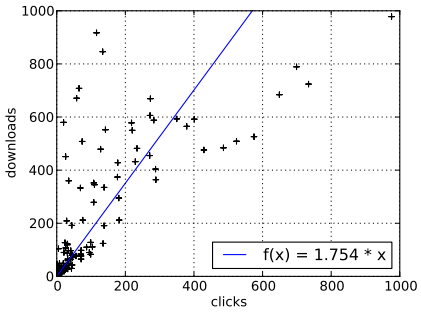
The median income for the top 50 highest earning uploaders is $ 11.74 for a median of 1.6 hours spent logged in and 10 files posted each day. (The top



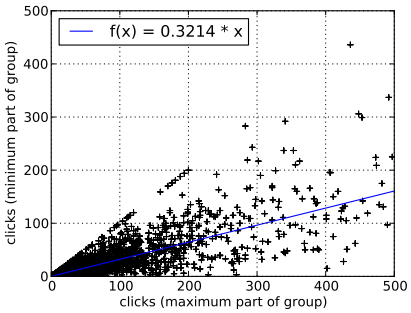
(a) `iload`: $\alpha \approx 0.40$ for VideoBB links.



(b) `redlist-mar`: $\alpha \approx 0.73$ for Files-Save.



(c) `redlist-mar`: $\alpha \approx 1.75$ for Fufox links.



(d) `redlist-mar`: $\alpha \approx 0.32$ for groups.

Fig. 7. Estimating the click-download ratio α by linear correlation: (a)–(c) use ground truth obtained from the OCH, (d) uses the difference between the most and least frequently clicked link in multipart downloads.

uploader earns $ 113.17 for an online time of around 8 hours and 200 files uploaded each day.) While this daily income would be worthwhile for an uploader based in a developing country, Fig. 9 shows that the vast majority of uploaders come from western countries, notably France. For reference, the current minimum legal wage in France is $ 12.50 per hour. This indicates that even the top uploaders earn relatively little compared to the work that they are doing.

Table 3(c) displays the overlap between the 50 uploaders with the highest income, links and clicks, respectively. Around 36 % of the users who provide most links are not among the best earning users. The fact that these uploaders do not imitate the behaviour of the best earning uploaders suggests that even the top uploaders do not all aim to maximise their income.

To assess the importance of the 50 highest earning uploaders for content availability on the site, we count how many content objects would become unavailable if all links provided by these users were removed. This corresponds to a scenario where the 50 highest earning uploaders stop uploading when

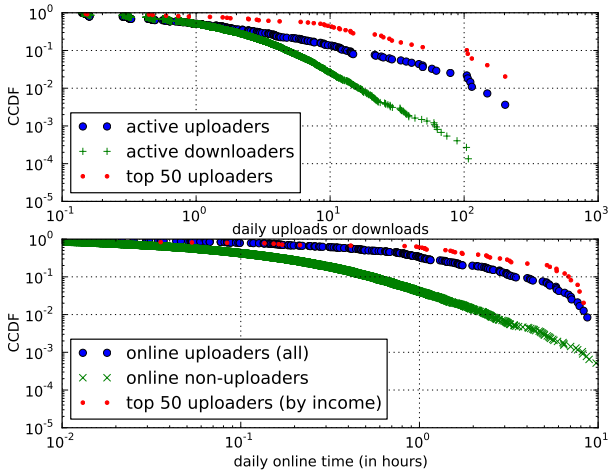


Fig. 8. User activity in `redlist-mar`: CCDF of the number of uploads and downloads per day (above) and the time spent logged in every day (below) for all users who logged in at least once during the crawl. For comparison, the plots include the 50 highest earning uploaders as an additional curve. Uploaders (and especially the top 50 uploaders) exhibit a more heavy-tailed activity than downloaders.

Table 3. Set intersections of the top 50 uploaders ranked by income (I), number of links (L) and number of clicks (C) in the two `redlist` data sets.

(a) between `oct11` and `mar12`

	Income	**Links**	**Clicks**
Income	14		
Links		7	
Clicks			5

(b) within `oct11`

	Links	**Clicks**
I	21	19
L		41

(c) within `mar12`

	Links	**Clicks**
I	32	42
L		37

the corresponding OCHs discontinue their affiliate programmes. We find that excluding the top 50 uploaders would remove 80 % of the total income and 58.5 % of all links, but only 39.7 % of the content objects and 21.7 % of the traffic: Many content objects have alternative download links provided by other users, and the content objects that have only links provided by the top uploaders are relatively unpopular overall. Consequently, anti-piracy measures aimed at disrupting economic upload incentives would have a limited effect on this site.

In summary, even most of the top 50 highest earning uploaders earn less than the minimum legal wage in their home country. Furthermore, `redlist` is rather resilient against the exclusion of its top 50 uploaders.

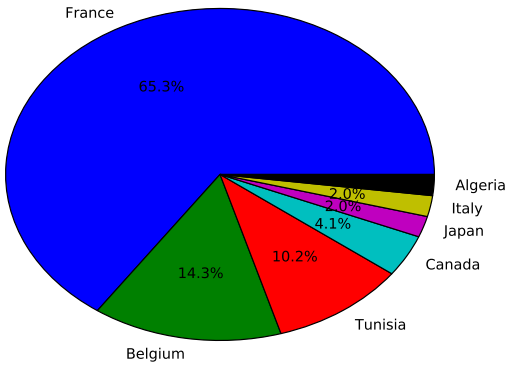


Fig. 9. Top 50 uploader's countries in `redlist-mar`: The vast majority of the highest earning uploaders come from western countries.

4.4 The Impact of the Megaupload Shutdown

In October 2011, before the shutdown, Megaupload was the most popular OCH on `redlist` as shown in Fig. 10(a). Because Megaupload had already ended its affiliate programme at that time, `redlist` generated just $ 184.79 per day.

After the shutdown of Megaupload, several other OCHs discontinued their affiliate programmes. Figure 10(b) shows that in March 2012, `redlist` used more OCHs than before that did not pay any rewards at all. Rapidshare, an OCH that had previously lost popularity due to its anti-piracy measures [4], regained significant popularity. Somewhat paradoxically, however, the shutdown of Megaupload lead to a more than fivefold increase in the daily income (up to $ 1,028.48) because Depositfiles and Uploaded, two OCHs with competitive affiliate programmes, became the two most popular OCHs on the site.

Overall, the number of available content objects decreased drastically by 62 % after Megaupload was closed, but the site quickly recovered and even increased its total click traffic by 5 %. These events illustrate that in the OCH ecosystem with its current diversity, even the shutdown of a major actor does not durably slow down the pace of file sharing.

5 Discussion

Our measurements show that the potential income of most uploaders is very low. Hence, these uploaders must have a different incentive rather than money. On the other hand, a few uploaders can earn significant amounts of money. This mix of uploader motivations has implications on proposed anti-piracy measures.

A recent focus in copyright enforcement appears to be on money flows [5]. For instance, the Stop Online Piracy Act (SOPA) proposed in the U.S. contains provisions to prevent advertising and payment services from processing payments in

relation to online piracy. While principally aimed at site operators, profit-oriented uploaders are indirectly affected by their dependence on affiliate programmes. However, as we have shown in this paper, profit-oriented uploaders are a small minority of all uploaders, and they are not essential for the ecosystem to survive. The majoritarian altruistic uploaders are not affected by this class of measures as long as sites remain available where they can upload and share their files.

More generally, our findings suggest that the overall impact of the OCHs' affiliate programmes on piracy activities may be overstated: Most users upload content despite earning next to nothing. Discontinuation of the affiliate programmes would deprive profit-oriented pirates of their illegal income, but it seems that these programmes are not the *main* driving force behind OCH-based piracy.

6 Related Work

Previous work in the area of OCH [1], [6], [9] focusses on network and workload-level measurements such as file sizes, download speeds, and the service architecture. These studies are partially based on network traces, and partially on crawls of indexing sites similar to our work. While some of the works depict OCH as an emerging alternative to BitTorrent for piracy, they do not deal with money-making opportunities or uploader motivations.

The closest work to ours is a short technical report published recently by Zubin Jelveh and Keith Ross [4]. The authors use payment screenshots posted in a webmaster forum to analyse the range of uploader income through Filesonic's PPD and PPS affiliate programme. In contrast to our work, the results by Jelveh and Ross reflect actual payouts. Based on 151 earnings screenshots covering 2,653 days, they report an average uploader income of $ 33.69 per day (minimum $ 0, maximum $ 226.27). This income range is generally consistent with what we find in our study. Beyond what we can analyse with our methodology, Jelveh and Ross find that income through PPD averages $ 21.12 as opposed to $ 46.10 through PPS. While providing actual ground truth data, the data set analysed by Jelveh and Ross suffers from a selection bias: Most uploaders do not make their income public. Furthermore, it is unknown where and how often uploaders post links, and what content they upload. While our methodology can only give an estimation of the actual uploader income, we compute a much more representative distribution of the income over uploaders. Furthermore, we use more comprehensive information about uploaders and content to calculate the value of links and the effort behind uploading, and we thereby obtain hints at the financial or altruistic motivations of uploaders.

Cuevas et al. [2] study the characteristics of initial seeders in BitTorrent. They find evidence for major initial content uploaders behaving in a non-altruistic way. Their result differs from our work in two ways: Firstly, BitTorrent does not have a direct mechanism to financially reward content uploaders; profit is usually generated by using uploads as a way to advertise external websites or to distribute malware. Secondly, it is common for OCH-based uploaders to copy a file and re-upload it on the same (or another) OCH. Therefore, OCH indexing

sites often have a high number of alternative downloads for the same content, which decreases the potential income for individual uploaders.

For a recent news article [5], Joe Karaganis conducted an anonymous income survey among BitTorrent site operators. Summarising the results, Karaganis characterises these sites as "financially fragile but low cost operations, dependent on volunteer labor, subsidized by users and founders, and characterized by a strong sense of mission to make work more widely available within fan communities".

7 Conclusion

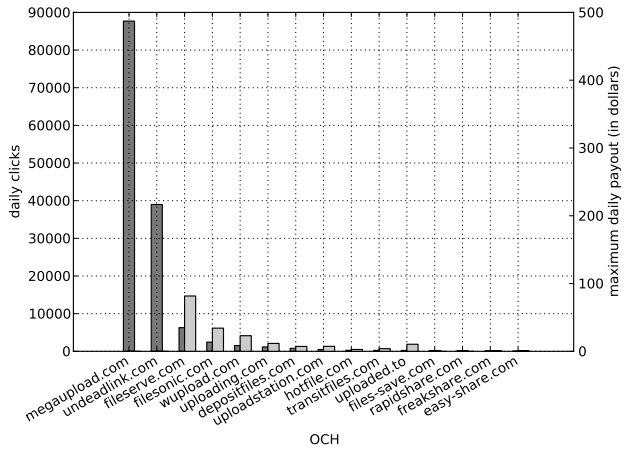
There is no black and white answer to uploader income in OCH-based file sharing: Most uploaders can earn only trivial amounts of money through OCHs' affiliate programmes and can be characterised as altruistic. A small number of very active uploaders, however, can earn in the order of hundred dollars per day and are more likely to be motivated by financial gain. Yet, the OCH file sharing ecosystem does not depend on these uploaders; most of the popular content would remain available if the links provided by the highest earning uploaders were excluded. The implication is that even in the OCH ecosystem with its money-mad reputation, anti-piracy measures that are premised on profit-driven uploader behaviour might not be as effective as the content industry believes. In order to sustainably address piracy, a holistic approach would be required that also removes incentives for altruistic uploaders, and for downloaders in general.

Acknowledgement

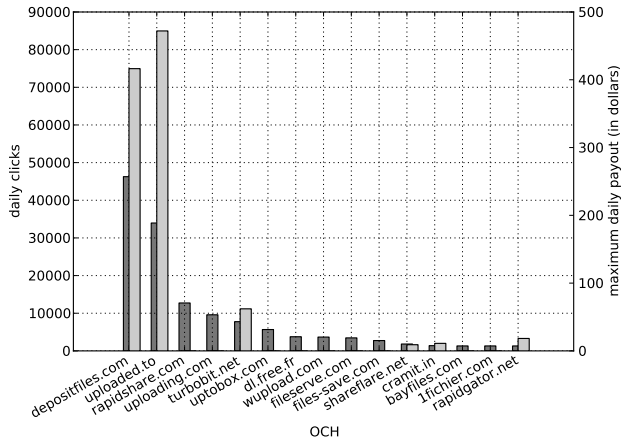
The authors thank Sy and Laurie Sternberg for their generous support.

References

1. Antoniades, D., Markatos, E., Dovrolis, C.: One-click hosting services: A file-sharing hideout. In: IMC '09 (Nov 2009)
2. Cuevas, R., Kryczka, M., Cuevas, A., Kaune, S., Guerrero, C., Rejaie, R.: Is content publishing in BitTorrent altruistic or profit-driven? In: Co-NEXT '10 (Nov 2010)
3. Howe, J.: The shadow Internet. Wired 13(01) (Jan 2005)
4. Jelveh, Z., Ross, K.: Profiting from piracy: A preliminary analysis of the economics of the cyberlocker ecosystem. Tech. rep., NYU-Poly, Brooklyn, USA (Jan 2012)
5. Karaganis, J.: Meganomics: The future of "follow-the-money" copyright enforcement. `http://torrentfreak.com/meganomics-the-future-of-follow-the-money-copyright-enforcement-120124/` (Jan 2012)
6. Mahanti, A., Williamson, C., Carlsson, N., Arlitt, M.: Characterizing the file hosting ecosystem: A view from the edge. In: Proc. IFIP PERFORMANCE (Oct 2011)
7. McCandless, D.: Warez world. Telepolis (Jul 2001)
8. Rehn, A.: The politics of contraband: The honor economies of the warez scene. Journal of Socio-Economics 33(3), 359–374 (2004)
9. Sanjuàs-Cuxart, J., Barlet-Ros, P., Solé-Pareta, J.: Measurement based analysis of one-click file hosting services. Journal of Network and Systems Management (May 2011)



(a) `redlist-oct`



(b) `redlist-mar`

Fig. 10. The daily number of clicks (dark grey) and daily maximum payout (light grey) for the Top 15 OCHs on `redlist` (a) before and (b) after the shutdown of Megaupload. Note that `dl.free.fr` is not shown in `redlist-oct` because the crawler did not recognise these links at that time.