Investigating the Effect of Anti-Piracy on the Popularity of P2P Applications

Angel Cuevas
Institut-Mines Télécom SudParis
angel.cuevas_rumin@it-sudparis.eu

Reza Farahbakhsh
Institut-Mines Télécom SudParis
reza.farahbakhsh@it-sudparis.eu

Ruben Cuevas
Universidad Carlos III de Madrid
rcuevas@it.uc3m.es

Michal Kryczka
Institute IMDEA Networks
michal.kryczka@imdea.org

Roberto Gonzalez
Universidad Carlos III de Madrid
rgonza1@it.uc3m.es

Reza Rejaie
University of Oregon
reza@cs.uoregon.edu

ABSTRACT

During the past decade, the growing popularity of Peer-to-Peer (P2P) applications has led to their significant impact on the Internet. The popularity of major P2P applications is primarily due to the availability of copyright-infringing content that is published by a small number of users (often for financial gain) and attracts a large number of consumers. This in turn has resulted in legal actions by copyright holders against P2P applications as well as new legislations and enforcement of anti-piracy laws. Despite the importance of these social and legal factors on the popularity of P2P applications, little is known about them.

This paper presents a measurement-based study that focuses on two problems: *(i)* Characterizing trends in the population of publishers and consumers along with their level of activity in the BitTorrent ecosystem, *(ii)* Investigating the effect of anti-piracy laws and their enforcement on the observed trends. In particular, our findings demonstrate that both country-specific laws and global events affect the behavior of participating users in illegal P2P applications. Well-publicized anti-piracy laws motivate consumers to reduce their level of activity even when the law does not directly target them. In contrast, major publishers adapt their operations to decrease the likelihood of any legal implication but continue their copyright infringement.

Keywords

BitTorrent, Measurements, Anti-Piracy, Megaupload

1. INTRODUCTION

The Internet has witnessed a dramatic growth in the popularity of Peer-to-Peer (P2P) applications (*e.g.*, BitTorrent, Gnutella) during the past decade. As a result, these popular P2P applications have produced a significant fraction of Internet traffic and thus profoundly affected the Internet and individual service providers (ISPs). For example, according to certain reports [1], more than 50% of Internet traffic was associated with BitTorrent in 2007. These trends in turn have led to a large and exciting body of work by researchers and practitioners on different aspects of P2P systems ranging from performance characterization [2, 3, 4, 5] and techniques for performance improvement [6, 7] to incorporating the P2P paradigm into new services.

However, the extreme popularity of P2P applications that generates their associated traffic and thus their impact on the network, seems to be primarily affected by social and economical (rather than technical) factors. More specifically, very popular P2P applications such as BitTorrent manage to attract a large number of users by making interesting copyrighted material (*e.g.*, recent release of Hollywood movies) freely available to them. Our recent investigation [8] revealed that a significant fraction of published copyrighted content in a major BitTorrent portal, namely The Pirate Bay, is provided by a small number of publishers. More importantly, we showed that these publishers commit this clear copyright infringement to attract the content consumers to their own web sites (or portals) for their own financial gain. The wide availability of copyrighted content by these P2P applications have resulted in legal actions by content owners against these P2P applications and their associated portals [9]. These events have also triggered the following two efforts that affects the popularity of P2P applications: *(i)* the growing attention to the need for new laws to deal with copyright issues on the Internet and its proper enforcements, and *(ii)* the emergence of cheap and copyright-compliant content delivery services over the Internet (*e.g.*, Netflix, Hulu) that might attract average users away from illegal P2P applications (*e.g.*, BitTorrent).

The economical, social and legal forces profoundly affect the availability of exciting content in P2P applications that fuels their popularity and thus their impact on the Internet. Despite the importance of these socio-economic factors, to our knowledge, their impacts on the popularity of P2P applications have not been examined and therefore are not well understood. For example, it is difficult to answer a basic question such as "how

do anti-piracy laws or competing copyright-compliant services affect the popularity of P2P applications among content publishers and consumers?". Clearly tackling these issues is challenging in large part due to the difficulties to identify, capture and characterize relevant socio-economic factors and their impact on P2P applications.

In this paper, we present a measurement-based study that tackles two problems:
(1) Characterizing the trends in the population and level of activity among publishers and consumers in one of the most important P2P applications, namely Bit-Torrent.
(2) Identifying a related set of socioeconomic factors and then investigating their impact on the observed trends.

We conduct an active measurement on the largest BitTorrent portal, The Pirate Bay, to capture the population of publishers and consumers as well as their level of activity in the BitTorrent ecosystem. Our dataset is formed by four snapshots of the BitTorrent ecosystem over a two years interval that overall consists of 120K torrents, 20K publishers, 55M consumers with more than 185M initiated download sessions. The first contribution of this study is the characterization of trends among publishers, consumers and their activities both at the aggregate level and at a per-country basis between Apr. 2010 and Feb. 2012. Our second contribution is an investigation that explores the effect of the following underlying socio-economic factors on the observed trends among publishers or consumers: *(i)* competing content delivery services, *(ii)* country-specific anti-piracy laws, and *(iii)* the closure of a major content sharing system, namely Megaupload [10]. In our investigation, we utilize the volume of search for relevant keywords from a certain location that is offered by Google trends [11] to assess the level of interest (or attention) among the corresponding population to specific applications or laws. While we may not be able to completely address an array of challenging questions that arise in our investigation due to our limited evidences and our resources, our study takes the first step to explore how socio-economic forces affect the popularity of P2P applications and thus their impact on the Internet.

Our main findings can be summarized as follows:
(1) The number of BitTorrent publishers, consumers and their activity have dropped between 2010 and 2012. Interestingly, the closure of Megaupload has speed up all these trends except the number of publishers that have dramatically increased.
(2) Exploring these trends at a per-country basis reveals that different countries may exhibit different deviations from the general trends or generally follow the same trend. The (dis)similarity of trends among countries for

a given period suggests whether its underlying causes are more likely to be local or global.
(3) The legal and illegal online Streaming services have became increasingly popular and have attracted the users away from BitTorrent.
(4) The implemented anti-piracy laws are more effective when the law and its enforcements are properly publicized and reach average users. Furthermore, the closure of Megaupload demonstrates that an anti-piracy action against a major player has a world-wide effect on the behaviour of content publishers and consumers that is more pronounced than local laws or events in individual countries.
(5) BitTorrent consumers and publishers react to anti-piracy laws and related enforcement events in a rather different manner. Average consumers conservatively reduce their level of activity in illegal sharing applications even when the law does not directly target consumers (*i.e.* they have not committed any illegal act). While regular publishers are among consumers and exhibit similar behavior, major publishers are not easily deterred by anti-piracy laws and related enforcement incidents. Major publishers appear to adapt their operation to reduce the likelihood of any legal implications but continue their copyright infringement.

The rest of this paper is organized as follows: Section 2 describes an overview of our data collection techniques and collected datasets. We characterize trends among BitTorrent publishers and consumers in Section 3 and 4, respectively. In Section 5, we present our investigation on the impact of the socio-economic root causes on the observed trends in the BitTorrent ecosystem. Section 6 offers a summary of related work and Section 7 concludes the paper.

2. DATA COLLECTION AND DATASETS

Our objective is to capture multiple snapshots of the BitTorrent ecosystem over time in order to characterize longitudinal trends in the population and activity of publishers and consumers. Towards this end, we leverage active measurement over The Pirate Bay (TPB) portal using the methodology and tools that we developed in our earlier study [8]. We focus on TPB in this study since it is one of the top-100 popular websites according to Alexa ranking [12] whose daily visits are at least twice (and in most cases significantly larger) than any other BitTorrent portal. Furthermore, all the indexed content on TPB portal are explicitly uploaded by a publisher in contrast to other major portals (*e.g.*, Torrentz or IsoHunt) that use crawling techniques to identify their indexed content. These features make TPB a suitable venue to capture snapshots of the Bit-Torrent ecosystem.

This section describes a brief overview on BitTorrent and our measurement methodology as well as main

Table 1: Datasets Description

	pb09	pb10	pb11	pb12
Crawling Period	11/28/09 - 12/18/09	04/09/10 - 05/05/10	10/21/11 - 12/13/11	01/28/12 - 02/12/12
Duration (days)	21	27	54	16
Pubishers (username)	3.8K	7.1K	6.9K	3.3K
Torrents	15.8K	38.2K	72.0K	21.0K
Consumers	-	27.3M	25.6M	5.1M
Downloads	-	95.6M	79.0M	11.1M

characteristics of our collected datasets.

Background on The Pirate Bay: TPB is simply a rendezvous point between content publishers and consumers. When a publisher wishes to make a content available within the BitTorrent ecosystem, its first step is to generate a unique id known as the *infohash* and register the content with one (or multiple) tracker(s). A tracker keeps track of the IP addresses for a group of peers that concurrently participate in the delivery of a content (*i.e.* form a swarm). A participating peer can be of two types: peers with a complete copy of a content are known as *seeders* while other peers are *leechers*. Therefore the content publisher is the first seeder in a swarm. The second step is to advertise the content by generating a .torrent file that provides meta-information for consumers including the IP address of the associated tracker(s). The publisher uploads the .torrent file to TPB and possibly other BitTorrent portals. In the case of TPB, the publisher needs to be registered with the portal and uses her account (with a specific username) to advertise a content. TPB creates a separate webpage for each registered user in which all its published content along with publishing times are listed. Finally, TPB offers an RSS service where consumers can subscribe and receive a notification as soon as a new content becomes available.

To download a content, a consumer typically retrieves the .torrent file from a portal, extracts the IP address of the tracker and connects to it. The tracker provides a list of IP addresses for a random subset of participating peers in the swarm to the new peer so that the new peer can connect to them and join the swarm.

An Overview of Measurement Methodology: Our measurement tool can capture a rather complete snapshot of all active publishers, their published files and associated consumers within a window of time. To achieve this goal, our tool subscribes to TPB's RSS service to get a notification for any new content that is published on the portal[1]. The RSS feed provides the .torrent file along with the username of the content publisher. Our tool retrieves the IP address of the tracker from the .torrent file (or the magnet link) and immediately connects to it. By connecting to the tracker right after the content is published, we are able to identify the IP address of the initial seeder (*i.e.* the publisher's location) in many torrents. Our tool periodically connects to the tracker to retrieve the IP addresses for (typically) 200 randomly-selected participating peers (*i.e.* consumers) while respecting the reconnection time imposed by the tracker in order to avoid being banned. To ensure that the IP address of most consumers are captured despite this limitation by the tracker, our tool probes a tracker from eight geographically-distributed nodes in parallel. Further details of our measurement methodology and its validations are described in our earlier work [8].

In summary, our measurement tool captures the following information for each published torrent on TPB portal: *(i)* publisher's username and IP address, *(ii)* list of IP addresses for associated consumers.

Datasets: Using our measurement tool, we have collected four snapshots of TPB system during the past 2 years. Table 1 summarizes the crawling period, the number of unique publishers, consumers, torrents (*i.e.* published files) and downloads for the four datasets labeled as pb09, pb10, pb11 and pb12. We note that pb09 dataset only includes information associated with publishers and thus it is only useful to characterize trends among publishers. Each dataset was collected over a sufficiently long time such that any common daily or even weekly variations among users and their activities are captured. Our consecutive snapshots have 4, 18 and 1.5 months gap between them where the last two snapshots are captured shortly before and after the closure of Megaupload site [10]. Therefore, we will use pb09, pb10 and pb11 to examine long-term trends while using pb11 and pb12 to investigate the effect of an important event, *i.e.* Megaupload closure.

3. PUBLISHER TRENDS

We start by characterizing the trends in the population of publishers and their level of publishing activities within TPB ecosystem as a representative of a broader BitTorrent ecosystem during our two year measurement period.

Metrics: Our snapshots of TPB users and activities have a different duration which makes it difficult to meaningfully compare their characteristics and identify a trend. To address this problem, we define the following two normalized metrics to capture the population of publishers and their activity independently of the duration of the measurement window for each snapshot:

---
[1]Note that since Feb. 2012, TPB only indexes magnet links instead of .torrent files. We have accordingly updated our tool to properly operate with this new indexing strategy.
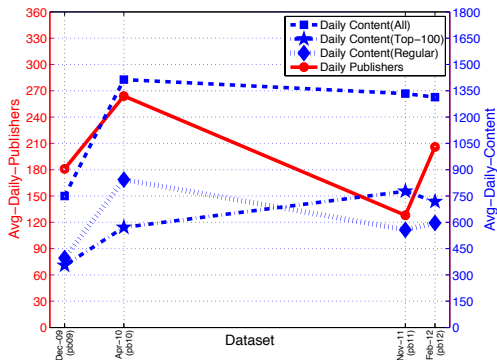
Figure 1: Publishing metrics (Avg-Daily-Publishers and Avg-Daily-Content) for the different datasets (pb09, pb10, pb11, pb12).

-Average Number of Daily Active Publishers (Avg-Daily-Publishers): This metric presents the normalized population of active publishers that exist on TPB each day of a given snapshot (*i.e. #publishers/day*).

-Average Number of Daily Published Content (Avg-Daily-Content): This metric indicates the daily level of publishing activity among publishers in TPB for each snapshot (*i.e. #contents/day*).

We can define these metrics across different group of publishers in each dataset to examine the trends among them. For example, one may calculate these measures over a subset of publishers who are located in a specific country and their corresponding published content to derive their trends. For all the presented analysis, we have examined the value of these measures across different windows of time to ensure that their value does not depend on the duration of their snapshot. We examine the main trends across different group of publishers to determine their behavior over time.

3.1 Aggregate Trends

We start by exploring the trends in the population and the level of activity across all publishers. Figure 1 shows these trends by depicting the evolution of our two metrics (each one in a separate y axis) across our four snapshots. The line labeled as "Daily Publishers" (referred to the left y axis) shows that the daily number of publishers increases by 45% between Dec. 2009 and Apr. 2010, then drops by 51% over the 18 months until Nov. 2011, and finally exhibits 60% increase in less than two months by Feb. 2012. Interestingly, the evolution of pubishers' activities, labeled as "Daily Content All" (referred to the right y axis), exhibits a different trend. It roughly doubles between Dec. 2009 and Apr. 2010, and then shows a steady and slow decrease of roughly 7% until Feb. 2012.

To gain more insight about the differences in the observed trends among publishers' population and activity, we take a closer look at the contribution of individ-
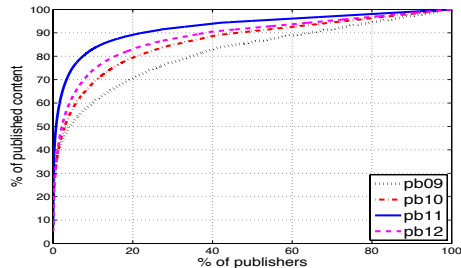

Figure 2: Skewness of published content (x axis - % of publishers, y axis - % of published content).

ual publishers. Figure 2 presents the CDF for the percentage of files that are published by the top x% of most contributing publishers in all four snapshots. This figure shows that the distribution of contributions among publishers is generally very skewed as a small fraction of publishers are responsible for a significant fraction of published files. However, Figure 2 reveals that the level of skewness in this distribution has changed over time. The contribution of top publishers has increased between pb09 and pb11 and then decreased in pb12. For example, the fraction of contributed content by the top 10% of all publishers across our four snapshots are 40%, 68%, 83% and 73%. The skewed nature of these CDFs suggests that it is useful to divide the publishers in each snapshots into two groups based on their contribution as follows: *(i) Top-100*: the top 100 publishers with the largest number of published files, and *(ii) Regular*: all other publishers that are ranked 100 and lower based on their contribution (all publishers except the top 100). Figure 1 shows the average daily number of published files by these two groups across our four snapshots. This figure reveals that the contribution of Top-100 publishers significantly increased from Nov. 2009 to Apr. 2010, and this trend continued until pb11 with a lower slope but reversed its direction after Nov. 2011. Publishing activities among regular publishers show an opposite trend to Top-100 publishers between Apr. 2010 and Jan. 2012. A closer examination of Top-100 publishers in pb11 and pb12 reveals that the drop in the publishing rate of Top-100 between Nov. 2011 and Jan. 2012 is primarily caused by the significant decrease in the publishing rate of 4 out of top 5 publishers, *i.e.* the decreasing trend is primarily driven by a handful of publishers.

3.2 Country-based Analysis

We now focus on the trends among publishers in individual countries since legal and social issues, which may affect the behavior of publishers, could vary across different countries. We use the IP address of each publisher to determine its geographical location (using Maxmind [13]) and then derive our two main metrics across all publishers that are mapped to each country. It must be noted that some publishers (13%, 24% and 23% for
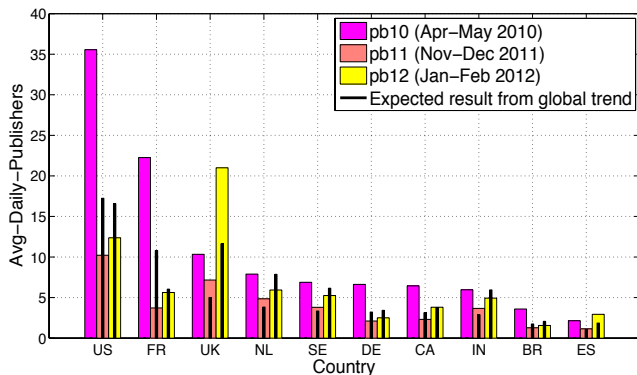
(a) Avg-Daily-Publishers
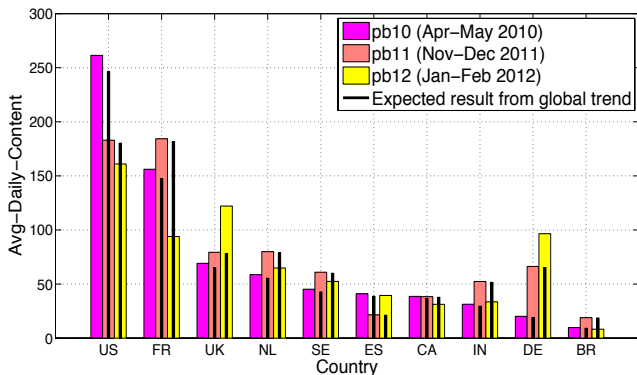
(b) Avg-Daily-Content

Figure 3: Analysis of publishing activity for the top countries in pb10, pb11 and pb12.

pb10, pb11 and pb12, respectively) upload content from multiple IPs in different countries. For these multi-country publishers we are able to accurately split the amount of content that they upload from each country.

Since we did not collect the IP address for publishers in pb09, we do not consider this dataset in this analysis. Figures 3(a) and 3(b) show the evolution of average daily population and average rate of publishing (*i.e.* our two metrics) across publishers in each country over three snapshots (pb10, pb11, pb12) with three adjacent bars. These figures present the results only for the top-10 countries with the largest publishing contribution[2].

We consider the observed pairwise changes in the daily number of publishers and their daily publishing rate between consecutive snapshots as a reference in this analysis. For example, if the daily number of publishers has the value of 100, 50, and 60 across pb10, pb11 and pb12, we multiply the measured daily number of publishers in pb10 (or pb11) by 0.5 (1.6) to derive its expected value in pb11 (pb12) if its step-wise changes would have been aligned with the aggregate trend of the corresponding metric. This is represented by the thin black line associated to the second and third bars in Figures 3(a) and 3(b). Using this pairwise changes as a reference for comparison ensures that the differences in behavior of aggregate and per-country trends do not propagate across multiple snapshots. We characterize the pairwise changes in the population and publishing rate of each country for two periods (pb10-pb11 and pb11-pb12) separately as follows:

Changes from pb10 to pb11: First, we observe that all countries decrease their number of daily publishers while all of them except for US[3] and ES also decrease their daily publishing rate. Furthermore, comparing the

value of the second bar and its corresponding line for each country in Figure 3(a) demonstrates that the number of publishers from US, FR, DE, CA and BR have dropped more than the aggregate trend whereas UK, NL and IN have added more publishers than average. This deviation from the aggregate behavior is particularly larger for US, FR and UK. Comparing the value of the second bar and its corresponding line for each country in Figure 3(b) shows that US and ES are the only countries whose publishers decrease the publishing rate more than the average rate while FR, UK, NL, SE, IN, DE and BR have increased their publishing rate more than the average trend. The countries that exhibit a larger gap from the average trend are US, FR, NL, ES and DE. Considering the overall trend in each country with respect to both metrics, US, UK, FR and ES are countries that globally deviate more from the average trend and deserve a major attention.

Changes from pb11 to pb12: First, we observe that all countries increase their number of daily publishers. Next, we focus on the third bar and its gap with the corresponding reference line for each country in both figures. Figure 3(a) reveals that the countries with significantly less number of publishers than the average trend in pb12 are US, NL while UK has added a much larger number of publishers than average in this period. As for published content, Figure 3(b) illustrates that UK, ES and DE show the largest positive gap while FR, US, BR exhibit the largest negative gap compared to the average trend for daily published content in this period.

4. CONSUMER TRENDS

In this section we turn our attention to trends among consumers in our snapshots. Since we did not capture the consumers in pb09, this analysis only focuses on other three snapshots. For this analysis, we follow similar steps as in the previous section.

Metrics: We define the following two normalized metrics to capture the population of consumers and their

---

[2]The selected 10 countries represent 40%, 31% and 32% of publishers, and 52%, 59% and 54% of the uploaded content for pb10, pb11 and pb12, respectively.

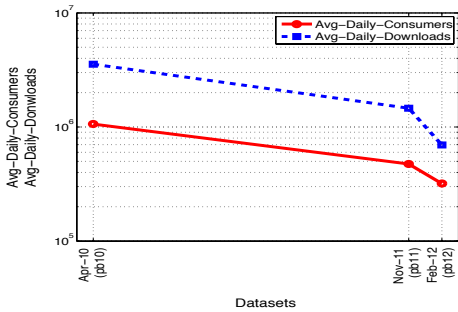[3]Note that we refer to countries using their 2-letters code defined by ISO 3166.

Figure 4: Consumers metrics (Avg-Daily-Consumers and Avg-Daily-Downloads) for the different datasets (pb10, pb11, pb12).

Figure 5: Skewness of consumed content (x axis - % of consumers, y axis - % of download sessions).

activity independently of the duration of the measurement window for each snapshot:

Average Number of Daily Active Consumers (Avg-Daily-Consumers): This metric presents the normalized population of active consumers per day for a given snapshot (*i.e. #consumers/day*).

Average Number of Daily Download Sessions (Avg-Daily-Downloads): This metric indicates the daily level of consuming (or downloading) activity among publishers for each snapshot (*i.e. #download sessions/day*).

4.1 Aggregate Trends

Figure 4 presents the average number of daily consumers and the average number of daily download sessions across the three snapshots using a log scale for y axis. This figure shows that during the 18 months between pb10 and pb11, both measures have dropped more than 50%. Both the population of consumers and their downloading activity continue to drop at an even faster pace during the second period. However, the decrease in the average number of downloads per day exhibits a larger drop (52%) compared to the average number of daily consumers (33%). Contrary to trends for publishers, trends for consumer-related metrics are correlated.

We examine the distribution of download rate among consumers in each snapshot to gain more insight. Figure 5 shows the percentage of download sessions (y axis) performed by the top x% of consumers (x axis) for pb10, pb11 and pb12. This figure shows that the contribution of consumers is generally skewed. The distribution looks very similar for pb10 and pb11 whereas the distribution becomes visibly less skewed for pb12. This suggests that the contribution of major consumers have clearly dropped within two months from pb11 to pb12.

4.2 Country-based Analysis

Similar to the presented analysis for publishers, we calculate our metrics across consumers and downloads associated with each country to obtain a country-specific view for trends among consumers and their downloading activities. Figures 6(a) and 6(b) present the aver-
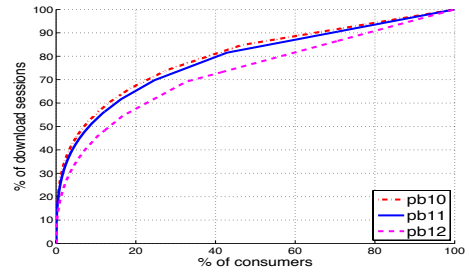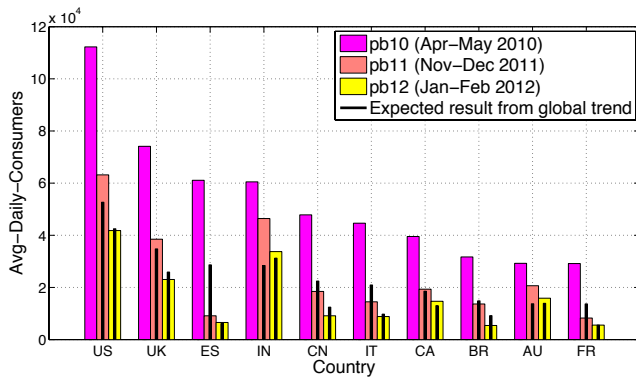
age number of daily consumers and the average number of download sessions per day for top-10 countries with largest consuming contributions[4]. Each figure shows the corresponding metric for three snapshots (pb10, pb11, pb12) with adjacent bars. We recall that the average number of daily consumers (and the average number of daily download sessions) dropped by 53% (59%) between pb10 and pb11, and further decreased by 33% (52%) between pb11 and pb12. The black thin line in each bar shows its expected value if that country would have changed with the same rate as the average trend across all consumers. These lines serve as a reference to compare changes of each metric between two consecutive snapshots with aggregate trends. It is interesting to note that only 7 countries are common between these two figures and the corresponding figures for publishers (in Figure 3). We examine the trends in both consumer-related metrics for each period (pb10-pb11, and pb11-pb12) as follows:

Changes from pb10 to pb11: First, we observe that all countries reduce their number of daily publishers and publishing rate. Next, comparing the bar related to pb11 with its corresponding line for each country in Figure 6(a) reveals that the number of consumers from US, UK, IN and AU is more than the aggregate trend whereas this number for ES, CN, IT and FR drops below the aggregate trend. Checking the trend in downloads between pb10 and pb11 in Figure 6(b) shows that countries with a larger than average number of consumers have a larger than average number of downloads (US, UK, IN, AU) and those with less than average number of consumers have less than average number of download (IT, ES, BR, FR, and CN). Overall, based on both metrics, US, ES, IT, and IN show largest deviation from the average trends in this period.

Changes from pb11 to pb12: Again, all countries reduce the number of daily publishers and publishing rate. Comparing the third bar (pb12) with its corresponding line for each country in Figures 6(a) and 6(b) shows that actual changes in most of the top-10 countries closely follow the aggregate trend. The only minor

---

[4]The selected 10 countries represent 53%, 53% and 52% of consumers, and 57%, 53% and 51% of download sessions for pb10, pb11 and pb12, respectively.

(a) Avg-Daily-Consumers
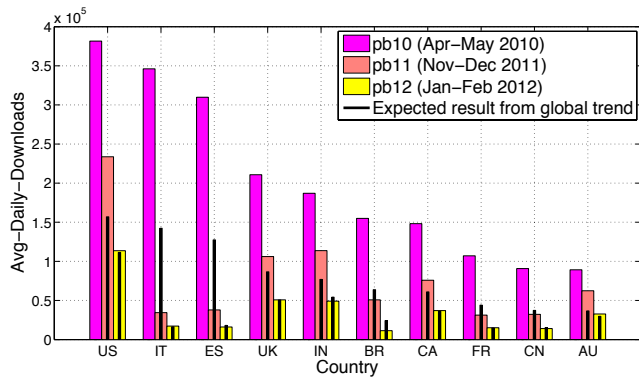
(b) Avg-Daily-Download

Figure 6: Analysis of consuming activity for the top countries in pb10, pb11 and pb12.

exception is BR where both the number of consumers and downloads dropped rather more than the average trend.

5. EXPLORING UNDERLYING CAUSES

Our goal is to identify the underlying (social, economical and legal) factors that derived the reported trends among publishers and consumers in previous sections. This is a challenging task at least for two reasons: First, it is very difficult to obtain all the relevant factors that may affect the number or the level of activity among publishers or consumers across our snapshots. Second, a specific trend might be affected by a combination of several factors. To address the first challenge, we use Google trends [11] to estimate the level of interest (or attention) to a certain content distribution system or a particular anti-piracy law at a global level or within a specific country. The provided trends for a searched keyword by Google show how the interest to that keyword has changed among Internet users over time. This service also allows one to obtain this information for searches that are performed by users in a specific country. To cope with the second challenge, we limit the scope of our investigation by exploring only the following three factors: *(i)* the other competing technologies for content distribution over the Internet, *(ii)* the country-specific anti-piracy laws, and *(iii)* the legal action against a major player (*i.e.* the closure of Megaupload). We examine the effect of the competing technologies and country-based anti-piracy laws on changes over the 18 months interval between pb10 and pb11. Then, the effect of Megaupload closure is examined on changes over 1.5 months between pb11 and pb12 that are captured shortly before and after this event.

5.1 Effect of Competing Systems (pb10-pb11)

There are broadly three major classes of content distribution systems that have been used over the Internet during the past few years as follows: *(i) P2P applications* like BitTorrent; *(ii) Cyberlockers* such as Megau-

pload in which the users directly retrieve their desired content from a server. Similar to BitTorrent and most P2P systems, Cyberlockers typically distribute copyrighted material and infringe copyright laws; *(iii) Streaming* systems directly "stream" a video or audio to a user which enables her to play the content as it arrives. Some of the streaming systems distribute content under agreements with content right holders such as Netflix or Spotify. There are also some streaming applications that distribute copyrighted content without the permission of the copyright holder, such as Megavideo (the streaming branch of Megaupload).

To assess how the relative popularity of these systems evolve over time between 2008 and 2012, we leverage the global Google trend for the following keywords that represent the above four alternative content distribution technologies:

- *"torrent download"* that represents a global view of BitTorrent which is the most popular P2P technology.
- *"Megaupload"* that before its closure was the most important Cyberlocker and one of the most popular Internet websites (*i.e.* according to Alexa [12], it was ranked among the 80 most popular websites in Jan. 2012).
- *"Megavideo"* as the most relevant copyright infringing streaming service ranked among the top 200 most popular websites based on Alexa Ranking.
- *"Netflix"* which is the major exponent of the copyright-compliant streaming services ranked among the top 100 websites around the world and #23 in US by Alexa. In addition, Netflix has been reported as the application responsible for the highest fraction of Internet traffic in North America [14].

Google trends for these keywords are shown in Figure 7 together with the time of data collection for our snapshots. Figure 7 demonstrates the following interesting points[5]:

---

[5]Clearly a single search for a keyword does not imply a single use of a particular system. For example, a user may access a particular system once she finds it via a Google search or a user may perform multiple searches and does not end
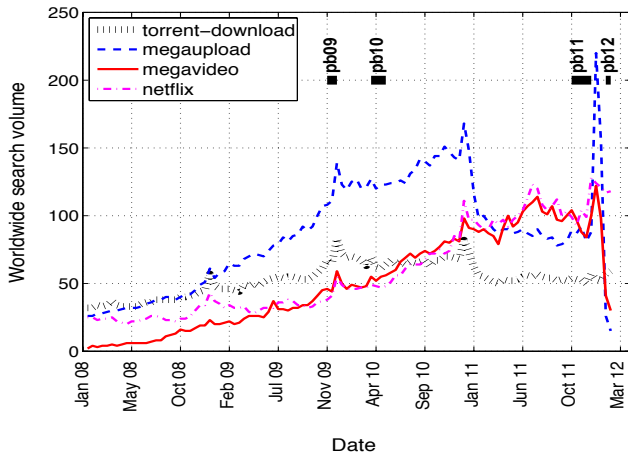
Figure 7: Competition analysis among content distribution technologies based of Google Trends search volume.

BitTorrent's popularity shows a steady growth until early 2010 when it becomes flat until early 2011 and then starts to decline. The popularity of Megaupload exhibits a bit different overall pattern than BitTorrent. It increases even faster than BitTorrent between 2008 and early 2010, and its growth slows down till early 2011 when it significantly drops in a short period and remains flat again until its closure in early 2012 (marked with a large spike). Megavideo had less popularity than previous two systems, but its popularity has steadily grown between 2008 and mid 2011 when it became flat and finally forced to close along with Megaupload in early 2012 (marked with a relatively large spike). Netflix's popularity was similar to BitTorrent and Megaupload in 2008 despite the fact that it was not a free service. Netflix's popularity has grown at a much slower pace than BitTorrent but has followed a steady pattern until early 2011. Despite the variations in its popularity during 2011 and 2012, Netflix continue to become more popular. Finally, it is worth noting that we always find spikes (for all the analyzed keywords) around the beginning of each new year. Although we do not have an evidence of why this happens, we founded it interesting to report this trend.

We assume that the popularity of these four systems represent the level of interest among Internet users

up using the corresponding system. Despite this subtle relationship between the volume of search for a keyword and the actual popularity of the related content distribution system, we believe that the volume of search offers a valuable and meaningful measure of interest among the average Internet users. Furthermore, we are not aware of any other scalable approach to accurately measure user interest to a particular system or a specific law in the society. A common approach for assessing user interest is through surveys that does not scale. Finally, we note that Google trends for proper keywords have been used to track the interest in other context such as spread of diseases [15].

(*i.e.* consumers[6]) to the corresponding class of content distribution systems. Given the steady growth in the number of users that access content over the Internet [14, 16, 17], reported trends in Figure 7 suggest that the consumers' interest has shifted away from BitTorrent to Cyberlockers and more visibly to Streaming services between Apr. 2010 (pb10) and Nov/Dec. 2011 (pb11). This trend is certainly aligned with the reported 53% drop in the population of consumers and 59% in the number of downloads between pb10 and pb11 snapshots. The dropping popularity of BitTorrent can be due to a combination of a few obvious reasons as follows: First, Cyberlockers offer an easier way for unskilled Internet users to download content (*i.e.* Google search and one click download) since they do not require installation of BitTorrent software. This behavior can be explained by the *Principle of least effort* [18] in Sociology. Second, Streaming services enables a user to view a content as it is being delivered which is more appealing than the required wait time in file sharing applications such as BitTorrent. Third, legal streaming services such as Netflix offer access to a large volume of content at a low price which is appealing to consumers who want to avoid any potential legal implications associated with downloading copyrighted content.

The drop in the number of BitTorrent consumers offers an explanation for observed trends in the number of published files by regular and major publishers. The regular publishers are often among consumers who exhibit altruistic behavior and share content. Therefore, the migration of consumers to other systems will affect the number of regular publishers and their interest. In contrast, the major publishers often publish copyrighted content to attract users to their own web site for financial gain as we demonstrated in our prior work [8]. These publishers reacted to the drop in consumers by increasing their publishing rate to attract more users and avoid any drop in the number of their consumers. This behavior is analogous to the common practice among companies to offer various incentives (*e.g.*, lower prices, more items or services) to stop the fly of customers [19].

5.2 Effect of Country-specific Anti-piracy Laws (pb10-pb11)

We now examine the effect of anti-piracy laws on the observed country-specific trends in the number and activity of BitTorrent publishers and consumers between pb10 and pb11. We focus on the country-specific trends simply because specific anti-piracy legislations are often passed and enforced at individual countries. We consider those top-10 countries whose specific trends more significantly deviated from the aggregate trends as we

---

[6]Note that the number of publishers is much smaller and they do not typically perform searches. Therefore, the observed trends primarily reflects the behavior of consumers.
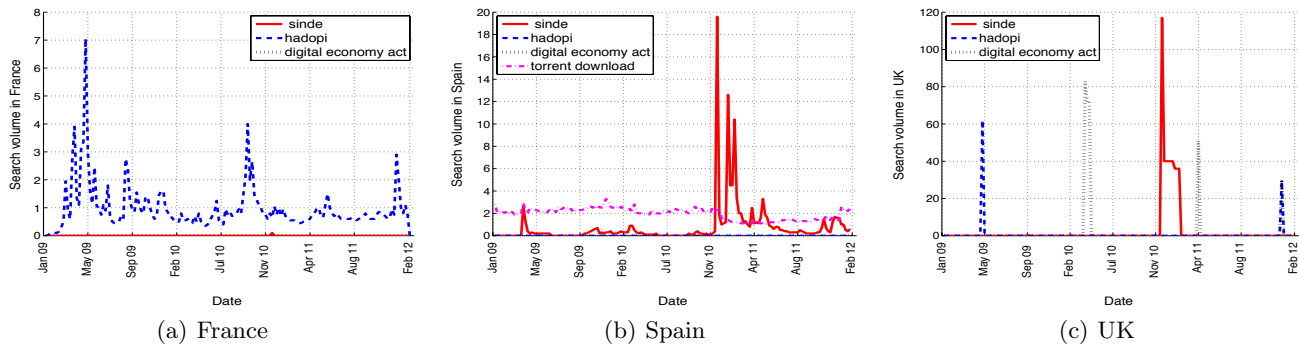
Figure 8: Popularity of the French, Spanish and British anti-piracy laws.

reported in Sections 3 and 4. Our intuition is that countries that exhibit trends similar to the aggregate trends, are affected by global factors such as competing technology rather than local factors such as anti-piracy laws. Among the list of top-10 countries that exhibited interesting trends, we focus our investigation on a handful of countries, namely USA (US), France (FR), United Kingdom (UK), Spain (ES) and Italy (IT). While there are clearly other countries among the top-10 that are worth such investigation, we need to limit the scope of this work to a few countries (with interesting trends) where information on their anti-piracy laws are readily available to us.

We briefly present the anti-piracy laws in each one of these countries, assess the level of interest (*i.e.* attention) among each country's population using Google trends, and then examine whether the collected information meaningfully explains the observed trends for that country between pb10 and pb11 in Sections 3 and 4. The three main anti-piracy laws in our analysis in this section are *Hadopi law* [20] in France, *Sinde law* [21] in Spain, and *Digital Economy Act* [22] in UK. Figure 8 shows the Google trend (search volumes) for the keywords "Hadopi", "Sinde" and "Digital Economy Act" in France, Spain and UK, respectively[7]. These figures reveal that the Sinde and in particular Hadopi laws received a fair amount of attention within their corresponding countries. Furthermore, Spanish and French citizens have paid a lot more attention to their own anti-piracy laws than the laws in their neighboring countries. In contrast, British showed very little attention to the Digital Economy Act which was comparable to their attention to the Hadopi and Sinde laws from neighboring countries. Next, we will discuss the trends in each country and relate them to their anti-piracy law and the level of attention among their citizens.

France: The anti-piracy law in France is called *Hadopi law*. This law was passed in Oct. 2009 and has been

[7]Note that the provided values in these Google trends are relative to the average search volume for one of the keywords from a specific geographical area. Therefore, we can only compare values within one figure while values across different figures are not comparable.

enforced since then. Hadopi is a three-strike law where an identified user downloading copyrighted material is warned twice and the third time her internet connection could be cut and/or she could be fined with up to 1500€. Interestingly, the Hadopi law exclusively targets users in P2P applications and does not affect users who download copyrighted content from Cyberlockers or Streaming systems. In the case of BitTorrent, both publisher and consumers in a swarm are punished by the Hadopi law. Figure 8(a) indicates that the Hadopi law has received plenty of attention within France between 2009 and 2012. The application of this law during 2011 have reported 470K first warning emails, around 20K a second warning, while only 10 subscribers received a third warning. Those 10 cases are now under court investigation [23].

In the past two sections, we showed that between pb10 and pb11, the number of publishers in France dropped 32% more than the aggregate trend while their number of published files was 24% larger than the aggregate trend. Also the number of consumers and their downloads dropped 19% and 12% more than the aggregate trend. This indicates that the Hadopi law has certainly managed to reduce both consumers and publishers more than the general trend. However, it is surprising that the number of published files have significantly increased despite the drop in the number of publishers while the Hadopi law was in effect. Our careful examination of publishers and their associated ISP led to an explanation for this unexpected trend. It turns out that a major hosting provider that has most of its data center in France, called OVH, provides hosting service for top BitTorrent publishers in France. These publishers, possibly from different countries, publish content into BitTorrent using the OVH service and do not seem to be concerned about the Hadopi law. In fact these publishers have increased their publishing rate and are primarily responsible for the 24% increase beyond the aggregate trend. This behavior suggests that the publishers at OVH do not feel threatened by the Hadopi law possibly due to the large size of OVH or their remote location outside of France.

Spain: Sinde law was approved by the Spanish parliament in Feb. 2011 and passed in Dec. 2011. The law attracted lots of attention from media and triggered an intense social debate during the first half of 2011. Figure 8(b) shows the Google trends for Sinde law and confirms this level of attention within Spain. The Sinde law aims at closing (or preventing the access) to those portals (and websites) that facilitate access to copyright infringing file sharing such as TPB.

In previous sections, we showed that between pb10 and pb11, the number of publishers from Spain was 5% more than aggregate trend while their rate of publishing was 42% less than the aggregate trend. After further investigation of Spanish publishers, we discovered that this significant reduction in publishing activity is mostly caused by the Top 5 publishers who were infringing copyright and at least three of them were associated with BitTorrent portals. More specifically, the most active publisher (*furtaperas*) interestingly migrated all its contribution to OVH (France), another 3 of them have left BitTorrent (*i.e.* were not present in pb11) and the last publisher dramatically reduced its activity. Therefore, it appears that the Sinde law has deterred these publishers. Our discovered trends for consumers showed that the number of consumers and downloads have dropped 32% and 29% more than the aggregate rate, respectively. Given the level of coverage for the Sinde law, we believe that this significant drop in the consumers' activity is in reaction to Sinde law. Despite the fact that the law does not target consumers at all, consumers are often unaware of these details and stopped using BitTorrent with the perception that it is illegal. To further verify this hypothesis, Figure 8(b) also compares the search volume associated to the keywords "Sinde" and "torrent download" from Spain as indicated by Google Trends. We observe a clearly visible decrease in the search volume associated to "torrent download" right after the first spike in search for Sinde law in late 2010. Finally, it is worth noting that at Mar. 1st 2012, 275 websites have been reported based on Sinde law, and those cases are currently under study [24].

Italy: While Italy did not present any anti-piracy law, it experienced a very significant anti-piracy event. In Feb. 2010, an Italian court forced Italian ISPs to block the access of their customers to TPB. Figure 9 presents the search volume for "the Pirate Bay" and "IsoHunt" (another major portal that index torrent files) from Italy to show the evolution of their popularity over time. We observe that after the court sentence the popularity of the Pirate Bay rapidly drops while Isohunt maintains its popularity. Therefore, the inability of average users to easily access TPB portal has rapidly and significantly reduced the number of consumers and their activity on this portal.
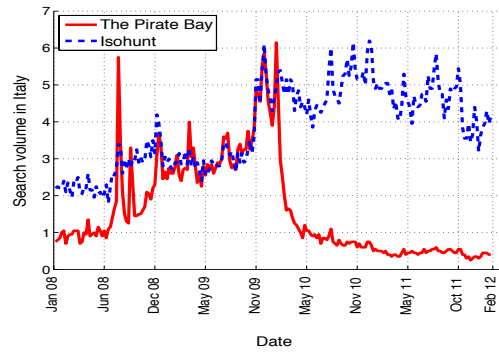

Figure 9: Popularity of The Pirate Bay and Iso-Hunt BitTorrent portals in Italy.

United Kingdom: UK's anti-piracy law, called Digital Economy Act (DEA) has been in effect since mid 2010. This law allows rights holders to gather lists of IP addresses for costumers that they believe to have infringed their copyrights and pass this information to the ISP. In essence, this law addresses both publishers and consumers. In earlier sections, we showed that between pb10 and pb11, the number of publishers and their published files were 21% and 20% larger than the aggregate trend, respectively. Similarly, the number of consumers and their downloads were 20% and 5% larger than the general trend. A key question is why the anti-piracy law in UK did not further reduce the participation of users in BitTorrent (*i.e.* it was not as effective as the laws in France and Spain)?

The main clue to answer this question is offered by Figure 8(c) that shows the search volume from UK for all three keywords associated with anti-piracy laws in UK, Spain and France. This figure reveals that British society paid attention to the DEA law only for a very short period of time around Nov. 2010. In fact, the level of interest to the DEA law was as small as their attention to the anti-piracy laws in the neighboring countries. This suggests that the lack of adequate attention by the British society has minimized the effect of the DEA law. Therefore, this law did not cause a significant deterrence to push the users away from BitTorrent.

US: Stop Online Piracy Act (SOPA) legislation was being considered in the US congress in late 2011 but it faced wide-spread protest by different Internet companies and did not pass. Therefore, there is no obvious legal force to deter publisher or consumers of copyright infringing file sharing systems. In prior sections, we showed that between pb10 and pb11, the US publishers have decreased their number and level of activity by roughly 20% and 24% more than the aggregate rate, respectively. During the same period, the number of consumers and their downloads have increased by 9% and 20% the average global trend, respectively. This suggests that the absence of any anti-piracy law did not provide any reason for publishers or consumers to leave BitTorrent at the aggregate rate.

5.3 Effect of Megaupload Closure (pb11 - pb12)

The Megaupload closure in Jan. 20th 2012 [10] attracted worldwide media coverage. The magnitude of this event is evident from the large spike in Figure 7 for the search volume associated with Megaupload (and Megavideo) around Jan 20th that is followed by a sharp drop. The closure of Megaupload was the main motivation for collecting pb12 snapshot shortly after this event. Since the time window between our pb11 and pb12 datasets is less than two months, and no other significant related event occurs during this window of time, any important change in BitTorrent popularity can be safely attributed to the closure of Megaupload.

Next, we examine the observed trends in Sections 3 and 4 between pb11 and pb12 and try to explain them with the Megaupload closure.

Migration of Megaupload Publishers to BitTorrent: We observed that the average number of publishers have increased 62% in two months. To explain this major change, we note that the closure of Megaupload disabled many copyright-infringing publishers from publishing their content. Most of these evicted publishers from Megaupload seem to have moved their activity to other platforms including BitTorrent. To verify this hypothesis and quantify the fraction of this newly-arriving publishers to BitTorrent, we have crawled TPB webpage associated to each one of the 3299 active publishers within our pb12 dataset. We obtain the date of the first published content on TPB for each publisher and consider that as the date that they joined this portal. Our analysis revealed that 42% of all present publishers on TPB in pb12, have published their first content after the closure of Megaupload. This is a good evidence that confirms the migration of these publishers from Megaupload and supports our hypothesis. Our analysis also demonstrated that these migrating publishers only contribute 20% of published content which in turn suggests that they are mostly regular publisher. Furthermore, the migration process is uneven across countries what leads them to present different deviations from the aggregate trend. Our careful examination showed that UK is the country accounting with the larger fraction of the newly arriving publishers to TPB (38%) and thus it is the country presenting the larger (positive) deviation from the average trend.

Finally, we observe that France presents an important reduction in the publishing rate below the aggregate trend. As we have seen, the overall content contribution from France comes mainly from publishers hosted by OVH. The reduction in the contributed content by French publishers in pb12 is due to the leave of some of the most important OVH publishers. However, we were unable to determine the reason for their sudden move.

Drop in the Activity of Top Publishers: Our results in Section 3 showed that 4 out of the 5 largest publishers in TPB drop their activity to half within two months from pb11 to pb12. This suggests that the legal action against Megaupload have caused concern among top publishers (that are publishing copyright-infringing content) and motivated them to reduce their level of activity to reduce the likelihood of possible legal actions against them. Such a behavior seems to be aligned with the theory in Economics [25, 26] that states that punishing a player who performs a non legitimate activity generates negative incentive on other players involved in similar activities.

Significant Drop in the Activity of Consumers: We have reported in Section 4 that the number of consumers and their activity dropped by 32% and 52% in pb12, respectively. Furthermore, our country-based analysis reveals that this significant drop in both consumer metrics is very similar across all countries (*i.e.* it is a world-wide phenomenon). These observations coupled with the lack of any other significant event in the considered period led to the conclusion that this drop in the consuming activity is caused by the Megaupload closure. Specifically, the wide coverage of Megaupload closure has caused concerns among the BitTorrent users who realize the similarity between these two systems and worried any legal action against them [25, 26].

Our main points in this section are the following:

(1) The growing popularity of competing technologies such as Cyberlockers and streaming services has affected the number of consumers in the BitTorrent ecosystem during the past couple of years.

(2) As we showed in the case of FR and ES, an antipiracy law would be more effective when the law and its enforcement are adequately publicized in the country.

(3) The closure of Megaupload shows that the antipiracy actions against major players have a significant world-wide effect on the behaviour of content publishers and consumers that is more pronounced than local laws or events in individual countries.

(4) The reaction of content consumers and publishers to anti-piracy laws and related enforcement events could be very different. Average consumers often conservatively reduce their level of activity in illegal file sharing applications in response to a well-publicized law or enforcement event even when these actions do not target consumers (as we observed for instance among Spanish consumers). However, major publishers with financial incentives tend to be more informed about legal issues and adjust their behavior (if needed) to minimize the likelihood of punishment associated to a specific law or enforcement event. For instance, publishers located at OVH in France do not reduce their contribution since the Hadopi law is not a direct threat for them whereas the biggest world-wide fishes halve their contribution after Megaupload closure due to their similar profile to avoid an eventual punishment.

6. RELATED WORK

There are several papers that look at the evolution of P2P traffic along the time *e.g.*, [27, 17]. The most recent one [17] studies the Inter-AS traffic associated to several ISPs across the Internet. The authors suggest that P2P traffic is becoming less representative and mention the migration process discussed in this paper as a possible cause. Furthermore, [28] studies the impact of BitTorrent in the Internet traffic over a period of two years between Nov 2008 and Nov 2010. The authors briefly mention a reduction of 10% in the number of peers that partially validates our observations. They argue that this reduction may be due to a drop in the system popularity and at the same time acknowledge the difficulty of validating this hypothesis so that they do not explore it. Our paper is different in nature than the previous works in the literature since we do not analyze the network footprint of BitTorrent, instead we perform a comprehensive analysis of the evolution of BitTorrent popularity at aggregate and local level across both publishers and consumers. In addition, we face the difficult task of finding the root causes for the discovered trends that to the best of our knowledge has not been addressed before.

7. CONCLUSION

In this paper we use data from 4 different BitTorrent snapshosts collected over a period of two years in order to characterize trends in the population of publishers and consumers and their respective activity levels. Furthermore, we explore the impact of different socio-economic forces into these trends. First, our results suggest that BitTorrent consumers have been migrating to other competing systems such as streaming services. Furthermore, we observe that consumers and publishers react in a different manner to laws and related enforcement events such as the closure of Megaupload. On the one hand, average consumers tend to reduce their download activity even when the law (or the event) does not aim to punish them. On the other hand, major publishers (typically motivated by financial incentives) seem to be more informed about specific legal implications of a law (or the enforcement event) and then adapt their activity (*e.g.*, decrease it), to avoid an eventual punishment, but rarely stop it. Finally, our results highlight that the effectiveness of any antipiracy effort (law or event) is dictated by the level of awareness among the population (*e.g.*, well-advertised laws are effective).

8. REFERENCES

[1] Ipoque, "Internet study 2007." [Online]. Available: http://www.ipoque.com/sites/default/files/mediafiles/documents/internet-study-2007.pdf

[2] L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, and X. Zhang, "Measurements, analysis, and modeling of bittorrent-like systems." in *Proc. of ACM IMC'05.*

[3] A. Legout, N. Liogkas, E. Kohler, and L. Zhang, "Clustering and sharing incentives in bittorrent systems," in *Proc. of ACM SIGMETRICS '07.*

[4] S. Kaune, R. Cuevas, G. Tyson, A. Mauthe, C. Guerrero, and R. Steinmetz, "Unraveling BitTorrent's File Unavailability: Measurements, Analysis," in *IEEE P2P'10*, 2010.

[5] C. Zhang, P. Dhungel, D. Wu, and K. Ross, "Unraveling the bittorrent ecosystem," *IEEE Transactions on Parallel and Distributed Systems*, 2010.

[6] M. Piatek, T. Isdal, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "Do incentives build robustness in BitTorrent?" in *Proc. of NSDI'07*, 2007.

[7] N. Laoutaris, D. Carra, and P. Michiardi, "Uplink allocation beyond choke/unchoke or how to divide and conquer best," in *Proc. of ACM CoNEXT'08.*

[8] R. Cuevas, M. Kryczka, A. Cuevas, S. Kaune, C. Guerrero, and R. Rejaie, "Is content publishing in bittorrent altruistic or profit-driven?" in *CoNEXT '10*, 2010.

[9] Wikipedia, "Legal issues with BitTorrent." [Online]. Available: http://en.wikipedia.org/wiki/Legal_issues_with_BitTorrent

[10] BBC, "Megaupload file-sharing site shut down." [Online]. Available: http://www.bbc.co.uk/news/technology-16642369

[11] Google, "Google trends." [Online]. Available: http://www.google.com/trends

[12] "Alexa." [Online]. Available: http://www.alexa.com

[13] "MaxMind- GeoIP." [Online]. Available: http://www.maxmind.com/app/ip-location

[14] Sandivine, "Sandvine global internet phenomena complete'," Spring 2011, Fall 2011, Spring 2012.

[15] H. A. Carneiro and E. Mylonakis, "Google trends: A web-based tool for real-time surveillance of disease outbreaks," *Clinical Infectious Diseases.*, vol. 49, no. 10, Oxford Journals. 2009.

[16] Cisco, "Visual Networking Index: Forecast and Methodology, 2011-2016," May 2012.

[17] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet inter-domain traffic," in *SIGCOMM '10*, 2010.

[18] G. K. Zipf, *Human Behaviour and the Principle of Least Effort: an Introduction to Human Ecology.* Addison-Wesley, 1949.

[19] Tefficient, "Free liberates france a second time?" [Online]. Available: http://www.tefficient.com/.cm4all/mediadb//tefficient%20public%20efficiency%20analysis%205.pdf

[20] Wikipedia, "Hadopi law." [Online]. Available: http://en.wikipedia.org/wiki/HADOPI_law

[21] Wikipedia., "Ley sinde." [Online]. Available: http://en.wikipedia.org/wiki/Ley_Sinde

[22] Wikipedia, "Digital economy act law." [Online]. Available: http://en.wikipedia.org/wiki/Digital_Economy_Act_2010

[23] TorrentFreak, "France tracks down 18 million file-sharers." [Online]. Available: ://torrentfreak.com/france-tracks-down-18-million-file-sharers-110714/

[24] El País (in Spanish), "El gobierno subraya el efecto disuasorio de la ley sinde." [Online]. Available: http://cultura.elpais.com/cultura/2012/03/23/actualidad/1332501881_052399.html

[25] G. S. Becker, "Crime and punishment: An economic approach," *Journal of Political Economy*, vol. 76, 1968.

[26] I. Ehrlich, "Crime, punishment, and the market for offenses," *Journal of Economic Perspectives*, vol. 10, no. 1, 1996.

[27] P. Borgnat, G. Dewaele, K. Fukuda, P. Abry, and K. Cho, "Seven years and one day: Sketching the evolution of internet traffic," in *INFOCOM*, 2009, pp. 711–719.

[28] J. S. Otto, M. A. Sánchez, D. R. Choffnes, F. E. Bustamante, and G. Siganos, "On blind mice and the elephant: understanding the network impact of a large distributed system," in *SIGCOMM '11 conference*, 2011.