This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Unveiling the Incentives for Content Publishing in Popular BitTorrent Portals

Rubén Cuevas, *Member, IEEE, ACM*, Michal Kryczka, Angel Cuevas, *Member, IEEE, ACM*, Sebastian Kaune, Carmen Guerrero, and Reza Rejaie, *Senior Member, IEEE, ACM*

***Abstract*—BitTorrent is the most popular peer-to-peer (P2P) content delivery application where individual users share various types of content with tens of thousands of other users. The growing popularity of BitTorrent is primarily due to the availability of valuable content without any cost for the consumers. However, apart from the required resources, publishing valuable (and often copyrighted) content has serious legal implications for the users who publish the material. This raises the question that whether (at least major) content publishers behave in an altruistic fashion or have other motives such as financial incentives. In this paper, we identify the content publishers of more than 55 K torrents in two major BitTorrent portals and examine their characteristics. We discover that around 100 publishers are responsible for publishing 67% of the content, which corresponds to 75% of the downloads. Our investigation reveals several key insights about major publishers. First, antipiracy agencies and malicious users publish "fake" files to protect copyrighted content and spread malware, respectively. Second, excluding the fake publishers, content publishing in major BitTorrent portals appears to be largely driven by companies that try to attract consumers to their own Web sites for financial gain. Finally, we demonstrate that profit-driven publishers attract more loyal consumers than altruistic top publishers, whereas the latter have a larger fraction of loyal consumers with a higher degree of loyalty than the former.**

***Index Terms*—BitTorrent, business model, content publishing, measurements.**

I. Introduction

PEER-TO-PEER (P2P) file-sharing applications, and more specifically BitTorrent, are clear examples of *killer* Internet applications of the last decade. BitTorrent is currently used by hundreds of millions of users and is responsible for a large portion of the Internet traffic [8]. This, in turn, has motivated the research community to examine various aspects of the swarming mechanism in BitTorrent [15], [18], [22] and propose different techniques to improve its performance [21], [25]. Furthermore, other aspects of BitTorrent such as the demography of users [17], [26], [28] along with security [27] and privacy issues [12], [13] have also been studied. However, the socioeconomic aspects of P2P file-sharing systems in general and BitTorrent in particular have received little attention. More specifically, a key factor in the popularity of BitTorrent is the availability of popular and often copyrighted content (e.g., recent TV shows and Hollywood movies) to millions of interested users at no cost. This raises an important question about the incentive of publishers who make these files available through BitTorrent portals. To our knowledge, prior studies on BitTorrent have not addressed this critical question.

In this paper, we study content publishing in BitTorrent from a socioeconomic point of view by unveiling *who* publishes content in major BitTorrent portals and *why*. Toward this end, we conduct a large-scale measurement over two major BitTorrent portals, namely Mininova [6] and the Pirate Bay [10], to capture more than 55 K published content objects that involve more than 35 M IP addresses. Using this dataset, we first examine the contribution of the individual content publishers and illustrate that around 100 publishers are responsible for uploading 67% of the published files that serve 75% of the unique peer downloads in our major dataset. Furthermore, most of these major publishers dedicate their resources for publishing content while consuming little to none content published by others, i.e., their level of content publication and consumption is very imbalanced. In addition to allocating a significant amount of resources for publishing content, these users often publish copyrighted material, which has legal implications for them [1], [2]. These observations raise the following question: *What are the main incentives of (major) content publishers in large BitTorrent portals?*

To answer this important question, we conduct a systematic study on major BitTorrent publishers. We show that these publishers can be broadly divided into two different groups: *fake publishers* who publish a large number of fake content and *top publishers* who publish a large number of often copyrighted content. We also identify the main characteristics (i.e., signature) of publishers in each group such as their seeding behavior and the popularity of their published content. We investigate the main incentives of major (non-fake) publishers and classify them into the following three categories (or profiles): 1) *private BitTorrent portals* that offer certain services and receive financial gain through ads, donations and fees; 2) *promoting Web sites* that leverage published content at BitTorrent portals to attract users to their own Web sites for financial gain; and

Manuscript received November 23, 2011; revised June 25, 2012; accepted October 10, 2012; approved by IEEE/ACM Transactions on Networking Editor T. Karagiannis. This work was supported in part by the European Union under the FP7 TREND(257740) Project, the Spanish Ministry of Economy and Competitiveness under the eeCONTENT Project (TEC2011-29688-C02-02), the Regional Government of Madrid under the MEDIANET Project (S-2009/TIC-1468), and the National Science Foundation under Grant IIS-0917381.

R. Cuevas and C. Guerrero are with the University Carlos III of Madrid, Leganés 28911, Spain (e-mail: rcuevas@it.uc3m.es).

M Kryczka is with the University Carlos III of Madrid, Leganes 28911, Spain, and also with the Institute IMDEA Networks, Madrid 28918, Spain.

A. Cuevas is with the Institut-Mines Telecom SudParis, Evry 91011, France.

S. Kaune is with the Multimedia Communications Lab (KOM), Technische Universität Darmstadt, Darmstadt 64283, Germany.

R. Rejaie is with the University of Oregon, Eugene, OR 97403 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNET.2012.2228224

3) *altruistic top publishers*. We characterize these three groups of publishers and estimate the value (and income) of their associated Web sites to support our claims about their incentives.

Finally, we define the notion of consumer loyalty toward a particular publisher and examine loyalty among BitTorrent consumers. We demonstrate that loyal consumers are mostly associated with top publishers. Furthermore, the fraction of loyal consumers and their level of loyalty toward a top publisher appears to be related to the publisher's profile as well as the type and the amount of published content.

The main contributions of this paper can be summarized as follows.

- We present a simple measurement methodology to monitor the content publishing activity in large BitTorrent portals. This methodology has been implemented in a system that continuously monitors and reports the content publishing activity in the Pirate Bay portal. The collected data by our system is made publicly available through our project Web site [3].
- The distribution of the number of published content by each publisher is very skewed, i.e., a very small fraction of publishers (3%) is responsible for a significant fraction of the published content (67%) and for an even more significant fraction of download sessions (75%). These major publishers can be further divided into three groups based on their incentives as follows: *fake* publishers, *altruistic top* publishers, and *profit-driven top* publishers.
- *Fake* publishers are either antipiracy agencies or malicious users who are responsible for 30% of the content and 25% of the downloads. These publishers sustain a continuous content poisoning attack [23] against major BitTorrent portals that affects millions of downloaders.
- *Profit-driven top* publishers own fairly profitable Web sites. They use major BitTorrent portals such as the Pirate Bay as a platform to attract millions of users to their Web sites by showing the associated URL to a user at different steps of the file download. The publishers that pursue this approach are responsible for roughly 26% of the content and 40% of the downloads in major BitTorrent portals.
- We show that only top publishers attract a significant number of loyal consumers. More interestingly, we observe that profit-driven top publishers attract a larger number of loyal consumers than altruistic top publishers (55 K versus 6 K), whereas altruistic top publishers have a larger fraction of loyal consumers (33% versus 13%) with a higher level of loyalty. This distinction appears to be directly related to their publishing strategy.

The rest of this paper is organized as follows. Section II describes our measurement methodology. Sections III and IV are dedicated to the identification of major publishers and their main characteristics (i.e., signature), respectively. In Section V, we study the key incentives for major content publishers. Consumer loyalty and its relationship with publisher profiles are examined in Section VI. Section VII presents other players that also benefit from content publishing. In Section VIII, we describe our publicly available application to monitor content publishing activity in the Pirate Bay portal. Finally, Section IX discusses related work, and Section X concludes the paper.

II. Measurement Methodology

This section describes our methodology to identify the initial publisher of a file distributed through a BitTorrent swarm. Toward this end, we first briefly describe the required background on how a user joins a BitTorrent swarm.

Background: A BitTorrent client takes the following steps to join the swarm associated with file X. First, the client obtains the .torrent file associated to the desired swarm. The .torrent file contains contact information for the tracker that manages the swarm and the number of pieces of file X. Second, the client connects to the tracker and obtains the following information: 1) the number of seeders and leechers that are currently connected to the swarm, and 2) N (typically 50) random IP addresses of participating peers in the swarm. Furthermore, if the number of neighbors is eventually lower than a given threshold (typically 20), the client contacts the tracker again to learn about other peers in the swarm.

To facilitate the bootstrapping process, the .torrent files are typically indexed at BitTorrent portals. Some of the major portals (e.g., the Pirate Bay or Mininova) index millions of .torrent files [28], classify them into different categories, and provide a Web page with detailed information (content category, publisher's username, file size, and file description) for each file. These portals also offer an RSS feed to announce every new published file. The RSS feed provides some information such as content category, content size, and publisher's username for a new file.

Identifying the Initial Publisher: The objective of our measurement study is to determine the identity of the initial publishers of a large number of torrents and to assess the popularity of each published file (i.e., the number and identity of peers who download the file).

Toward this end, we leverage the RSS feed to detect the availability of a new file on major BitTorrent portals and retrieve the publisher's username. In order to obtain the publisher's IP address, we immediately download the .torrent file and connect to the associated tracker. This implies that we often contact the tracker shortly after the birth of the associated swarm when the number of participating peers is likely to be small and includes the initial publisher (i.e., seeder). We retrieve the IP address of all participating peers as well as the current number of seeders in the swarm. If there is only one seeder in the swarm and the number of participating peers is not too large (i.e., <20), we obtain the bitfield of available pieces at individual peers to identify the seeder. Otherwise, reliably identifying the initial seeder is difficult because either there are more than one seeder or the number of participating peers is large.[1] Furthermore, we cannot directly contact the initial seeder that is behind a NAT box, and thus we are unable to identify the initial publisher's IP address in such cases. Using the above techniques, we were able to reliably identify the publisher's username for all the torrents and the publisher's IP address in at least 40% of the torrents.

[1]Our investigation revealed two interesting scenarios for which we could not identify the initial publisher's IP address: 1) swarms that have a large number of peers shortly after they are added to the portal; we discovered that these swarms have already been published in other portals; 2) swarms for which the tracker did not report any seeder for a while or did not report a seeder at all.

TABLE I
DATASETS DESCRIPTION: PORTAL NAME, START AND END DATES, NUMBER OF COLLECTED TORRENTS FOR WHICH WE IDENTIFY THE USERNAME AND THE IP ADDRESS OF THE INITIAL SEEDER, AND NUMBER OF COLLECTED CONSUMERS' IP ADDRESSES

	Portal Name	Start Date	End Date	#Torrents (username/IP)	#IP addresses
mn08	Mininova	09-Dec-08	16-Jan-09	- /20.8K	8.2M
pb09	Pirate Bay	28-Nov-09	18-Dec-09	23.2K/10.4K	52.9K
pb10	Pirate Bay	06-Apr-10	05-May-10	38.4K/14.6K	27.3M

Once we identify a publisher, we periodically query the tracker in order to obtain the IP addresses of the participants in the associated swarm and always solicit the maximum number of IP addresses (i.e., 200) from the tracker. To avoid being blacklisted by the tracker, we issue our queries at the maximum rate allowed by the tracker (i.e., 1 query every 10–15 min depending on the tracker load). Given this constraint, we query the tracker from eight geographically distributed machines[2] so that all these machines collectively provide an adequately large probing rate to quickly discover most (and often all) of the participating peers and their evolution over time. We continue to monitor a target swarm until we receive 10 consecutive empty replies from the tracker. We use the MaxMind Database [5] to map all the IP addresses (for both publishers and downloaders) to their corresponding ISPs and geographical locations.

A. Dataset

Using the described methodology, we identify a large number of BitTorrent swarms at two major BitTorrent portals, namely Mininova and the Pirate Bay. Each one of these portals was the most popular BitTorrent portal at the time of the corresponding measurement according to Alexa ranking [4]. It is worth noting that the Pirate Bay is particularly interesting for our study since it is the only main BitTorrent portal where all the published content is contributed by users [28] (as opposed to being retrieved from other portals). Table I shows the main features of our three datasets (one from Mininova and two from the Pirate Bay) including the start and end dates of our measurement, the number of torrents for which we identified the initial publisher (username/IP address), and the total number of discovered IP addresses associated for all the monitored swarms. We refer to these datasets as *mn08*, *pb09*, and *pb10* throughout this paper. We note that dataset *mn08* does not contain the username of initial publishers, and for dataset *pb09*, we use a single query to identify initial publishers after detecting a new swarm through the RSS feed, but do not probe the tracker to capture all the consumers.

III. IDENTIFYING MAJOR PUBLISHERS

A publisher can be identified by its username and/or IP address. In our analysis, we identify individual publishers primarily by their username since the username is expected to remain consistent across different torrents. However, we require the identification of an individual publisher's IP address for network-level analyses such as determining the ISP where a publisher is located, the duration of time that a publisher remains in a swarm, or its participation across multiple swarms either as a publisher or a consumer. For these network-level analyses, we only consider those torrents in out dataset for which we are able to identify the IP address of the publisher (i.e., the initial seeder). Furthermore, we can identify publishers in *mn08* dataset only by their IP addresses since this dataset does not include the username of individual publishers. Finally, *fake* publishers are better identified by their IP addresses as we describe in the following.

[2]One located in Oslo, Norway; one in Barcelona, Spain; one in Albacete, Spain; and five in different locations in Madrid, Spain.

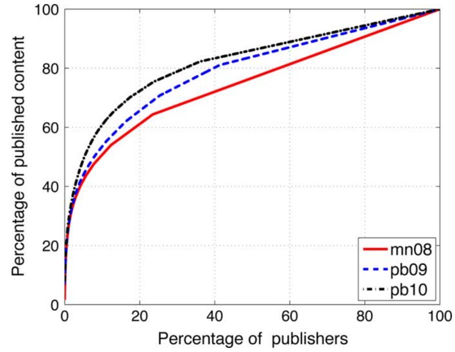


Fig. 1. Percentage of content published by the top x% publishers.

A. Skewness of Contribution

First, we examine the level of contribution (i.e., the number of published files) by the identified content publishers in each dataset. Fig. 1 depicts the percentage of files that are published by the top x% of the publishers in our three datasets. We observe that the top 3% of the BitTorrent publishers contribute roughly 40% of the published content. Moreover, a more careful examination of IP addresses for the top-100 (i.e., 3%) publishers in our *pb10* dataset reveals that a significant fraction of them either do not download any content (40%) or download less than five files (80%). This large contribution of resources (bandwidth and content) by major publishers coupled with the significant imbalance between their publishing and consumption rates appears nonaltruistic and rather difficult to justify for two simple reasons.

- *Required resources/cost:* Publishing a large number of content requires a significant amount of resources (e.g., bandwidth). For example, a major content publisher named *eztv* recommends in its private BitTorrent portal Web page (www.eztv.it) to allocate at least 10 Mb/s in order to sustain the seeding of few (around five) files in parallel.
- *Legal implications:* As other studies have reported [12] and we confirm in our datasets, a large fraction of content published by major publishers is copyrighted material (e.g., recent movies or TV series). Thus, publishing these files is likely to have serious legal consequences for these publishers [1], [2].

This raises the following question: *Why do a small fraction of publishers allocate a great deal of (costly) resources to contribute many files into BitTorrent swarms despite potential legal implications?* We answer this central question in Section V.

B. Publishers' ISPs

To help identify content publishers in our dataset, we determine the ISP that hosts each major publisher and use that

TABLE II
CONTENT PUBLISHERS DISTRIBUTION PER ISP

mn08			pb09			pb10		
ISP	Type	%	ISP	Type	%	ISP	Type	%
OVH	Hosting Provider	13.31	OVH	Hosting Provider	24.76	OVH	Hosting Provider	15.16
Comcast	Commercial ISP	4.69	Comcast	Commercial ISP	3.67	SoftLayer Tech.	Hosting Provider	4.52
Keyweb	Hosting Provider	3.18	Road Runner	Commercial ISP	2.3	FDCservers	Hosting Provider	3.64
Road Runner	Commercial ISP	3.03	Romania DS	Commercial ISP	2.27	Open Computer Network	Commercial ISP	3.59
NetDirect	Hosting Provider	2.44	MTT Network	Commercial ISP	1.95	tzulo	Hosting Provider	3.36
Virgin Media	Commercial ISP	2.42	Verizon	Commercial ISP	1.64	Comcast	Commercial ISP	2.86
NetWork Operations Center	Hosting Provider	2.39	Virgin Media	Commercial ISP	1.49	Cosema	Commercial ISP	2.25
SBC	Commercial ISP	2.38	SBC	Commercial ISP	1.41	Telefonica	Commercial ISP	2.22
Comcor-TV	Commercial ISP	2.33	NIB	Commercial ISP	1.26	Jazz Telecom.	Commercial ISP	2.07
Telecom Italia	Commercial ISP	2.02	tzulo	Hosting Provider	1.14	4RWEB	Hosting Provider	2.06

information to assess the type of service (and available resources) that a publisher is likely to have. Toward this end, we map the IP address for a publisher in each dataset to its corresponding ISP using the MaxMind database [5]. We then examine the publicly available information about each ISP (e.g., its Web page) to determine whether it is a commercial ISP or a hosting provider. We perform this analysis only for the top-100 (roughly 3%) publishers since these publishers are mostly of interest and collecting the required information for all publishers is a tedious task. Since we do not have publishers' username for *mn08*, we examine the top-100 publishers based on their IP addresses in this dataset. For these publishers, we cannot assess the aggregated contribution of a publisher through different IP addresses (i.e., under-estimating the contribution of each publisher).

We observe that 42% of the top-100 publishers in *pb10*, 35% of the top-100 in *pb09* and 77% of the top-100 publishers in *mn08* are located at hosting services. Moreover, 22%, 20%, and 45% of these top-100 publishers are located at a particular hosting services, namely OVH, in *pb10*, *pb09*, and *mn08*, respectively.

In short, our analysis reveals that a significant fraction of major publishers are located at a few hosting services and a large percentage of them at OVH.

We also examine the contribution of BitTorrent publishers at the ISP-level by mapping all the publishers to their ISPs and identify the top-10 ISPs based on their aggregate published content for each dataset as shown in Table II. This table confirms that content publishers who are located at a particular hosting provider, namely OVH, have consistently contributed to a significant fraction of published content at major BitTorrent portals. There are also several commercial ISPs (e.g., Comcast) in Table II with a much smaller contribution.

To assess the difference between users from hosting providers and commercial ISPs, we compare and contrast the characteristics of all publishers that are located at OVH and Comcast as representative ISPs for each class of publishers in Table III. This table demonstrates the following two important differences: first, the aggregate contribution of each publisher at OVH is on average a few times larger than Comcast publishers. Second, Comcast publishers are sparsely scattered across many /16 IP prefixes and many geographical locations in the US, whereas OVH publishers are concentrated in a few /16 IP prefixes and a handful of different locations in Europe (i.e., the location of OVH's data centers). In essence, the content published by Comcast publishers comes from a large number

TABLE III
CHARACTERISTICS OF ALL OVH AND COMCAST PUBLISHERS IN *mn08, pb09,* AND *pb10*

	Published torrents	# IP addresses	# /16 IP Prefixes	# Geographical Locations
OVH (mn08)	2766	164	5	2
Comcast (mn08)	976	675	269	400
OVH (pb09)	2577	78	5	2
Comcast (pb09)	382	198	143	129
OVH (pb10)	2213	92	7	4
Comcast (pb10)	408	185	139	147

of typically altruistic users where each one publishes a small number of files likely from their home or work. In contrast, OVH publishers appear to be paying for a well-provisioned service to be able to publish a much larger number of files. We have also examined consuming peers (i.e., downloaders) in captured torrents and did not observe the presence of OVH users among these consuming peers.

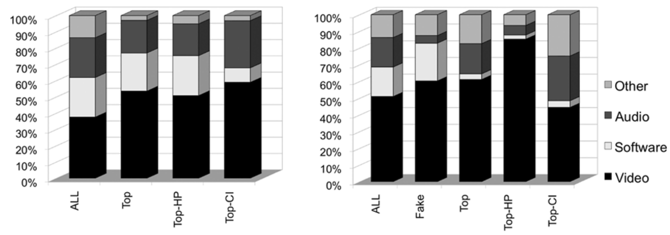
In summary, the examination of ISPs that host major BitTorrent publishers suggests that major publishers are located either at a few hosting providers (with a large concentration at OVH) or at commercial ISPs. These publishers contribute a significantly larger number of files than average publishers. Furthermore, publishers who are located at hosting providers do not consume published content by other publishers.

C. Closer Look at Major Publishers

We now examine the mapping between username and IP address of the top-100 content publishers in the *pb10* dataset to gain some insight about major publishers behavior. Our examination reveals the following interesting points.

First, if we focus on the top-100 IP addresses that have published the largest number of files, only 55% of them are used by a unique username. The remaining 45% of the IP addresses of major publishers are mapped to a large number of usernames. We have carefully investigated this set of IP addresses and discovered that they use either hacked or manually created accounts (with a random username) to inject "fake" content. These publishers appear to be associated with antipiracy agencies or malicious users. The former group tries to avoid the distribution of copyrighted content, whereas the latter attempts to disseminate malware. We refer to these publishers as *fake publishers*. Surprisingly, fake publishers are responsible for around 25% of the usernames, 30% of the published content, and 25% of the downloads in our *pb10* dataset. This suggests that major BitTorrent portals are suffering from a systematic poisoning

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) (b)

Fig. 2. Type of published content distribution for the different classes of publishers: *All*, *Fake*, *Top*, *Top-HP*, and *Top-CI*. (a) *mn08*. (b) *pb10*.

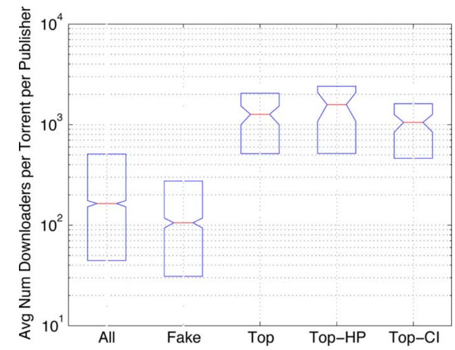


Fig. 3. Average number of downloaders per torrent per publisher for the different classes of publishers: *All*, *Fake*, *Top*, *Top-HP*, *Top-CI*.

index attack [23] that affects 30% of the published content. The portals fight this phenomenon by removing the fake content as well as the user accounts used to publish them. However, contrary to what has been reported in previous studies [26], this technique does not seem to be sufficiently effective since millions of users initiate the download of fake content. Finally, it is worth noting that most of the fake publishers perform their activity from three specific hosting providers named *tzulo*, *FDC Servers*, and *4RWEB*. Due to the relevant activity of these fake publishers, we study them as a separate group in the rest of the paper.

Second, the inspection of the top-100 usernames who publish the largest number of files shows that only 25% of them operate from a single IP address. The remaining 75% of top usernames utilize multiple IP addresses and can be classified into the following common cases:

1) 34% of the usernames with multiple IP addresses (5.7 IP addresses on average) located at a hosting provider in order to obtain the required resources for seeding a large number of files;
2) 24% of the usernames with multiple IP addresses (13.8 IP addresses on average) located at a single commercial ISP. Their mapping to multiple IP addresses could be due to the periodical change of their assigned IP addresses by their ISPs;
3) the other 17% of these usernames are mapped to multiple IP addresses (7.7 IP addresses on average) at different commercial ISPs. These are users who inject content from various locations (for example, a user may publish from both home and work computers).

To properly characterize the different types of publishers, we exclude the 16 usernames who publish fake content from the top-100 usernames. We refer to the remaining top-100 usernames (non-fake publishers) as *top* publishers who are responsible for 37% of the published content and 50% of the total downloads in our *pb10* dataset.

In summary, a significant fraction of the content is published by two group of publishers: top *publishers and* fake *publishers that are collectively responsible for 67% of the published content and 75% of the downloads*. In the rest of this paper, we devote our effort to characterize these two groups.

IV. Signature of Major Publishers

Before we investigate the incentives of major BitTorrent publishers, we examine whether they exhibit any other distinguishing features, i.e., whether major publishers have a distinguishing signature. Any such distinguishing features could shed some light on the underlying incentives of these publishers. Toward this end, in the following, we examine the following characteristics of major publishers in our datasets: 1) the type of published content; 2) the popularity of published content; and 3) the availability and seeding behavior of a publisher.

To identify distinguishing features, we examine the above characteristics across the following three *target groups* in each dataset: all publishers (labeled as "All"), all fake publishers (labeled as "Fake"), and all top-100 (non-fake) publishers regardless of their ISPs (labeled as "Top"). We also examine the breakdown of top publishers based on their ISPs into hosting providers and commercial ISPs, labeled as "Top-HP" and "Top-CI," respectively.

A. Content Type

We leverage the reported content type by each publisher to classify the published content across different target groups. Fig. 2 depicts the breakdown of published content across different content type for all publishers in each target group for our Mininova and our major Pirate Bay datasets. We recall that without username information for each publisher in *mn08* dataset, we cannot identify fake publishers. Fig. 2 reveals a few interesting trends.

First, video files (which mainly include movies, TV shows, and pornographic content) constitute a significant fraction of published files across most groups with some important differences. The percentage of published video across all publishers is around 37%–51%, but it is slightly larger among top publishers. However, video is clearly a larger fraction of published content by the top publishers located at hosting providers in our *pb10* dataset. Fake publishers primarily focus on videos (recent movies and TV shows) and software content. This supports our earlier observation that these publishers consist of antipiracy agencies and malicious users because the former group publishes a fake version of recent movies, while the latter provides software that contains malware.

B. Content Popularity

The number of published files by a publisher shows only one dimension of its contribution to the BitTorrent ecosystem. The other equally important issue is the popularity of each published content (i.e., the number of downloaders regardless

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

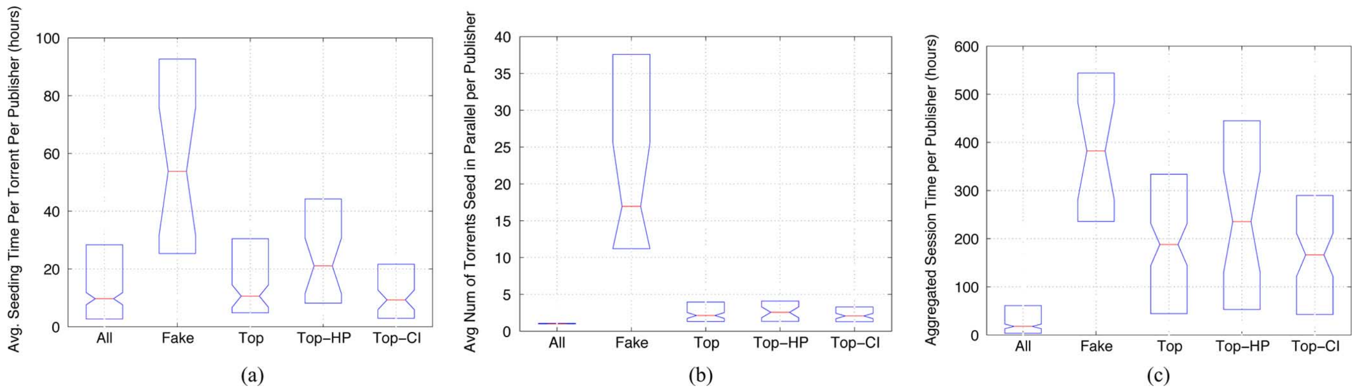


Fig. 4. Seeding behavior for the different classes of publishers: *All*, *Fake*, *Top*, *Top-HP*, and *Top-CI*. (a) Average seeding time per content per publisher. (b) Average number of parallel seed torrents per publisher. (c) Average aggregated session time per publisher.

of their download progress) by individual publishers. Fig. 3 shows the box plot of the distribution of average number of downloaders per torrent per publisher across all publishers in each target group, where each box presents the 25th, 50th, and 75th percentiles. All the box plots presented in the rest of this paper follow the same format.

On the one hand, the median popularity of top publishers' torrents is seven times higher than a typical user (represented by *All*). A closer examination of the top publishers shows that the content published by users at hosting providers is on average 1.5 times more popular than those published by users at commercial ISPs. On the other hand, fake publishers' content is the most unpopular among the target groups. This is because the portals actively monitor the torrents and immediately remove the content identified as fake to avoid users from downloading it. Furthermore, users quickly realize the fake nature of these content and report this information on forums that inform others and limit their popularity.

In summary, top publishers are responsible for a larger fraction of popular torrents. This, in turn, magnifies the contribution of the 37% of the injected files by the top publishers to be responsible for 50% of all the downloads. The low popularity of fake publishers' content has the opposite effect and limits their contribution to the number of downloads to 25%.

C. Seeding Behavior

We characterize the seeding behavior of individual publishers in our target groups using the following metrics: 1) average seeding time of a publisher for each published content; 2) average number of parallel seeded torrents; and 3) aggregated session time of a publisher across all its torrents. Since calculating these properties requires a detailed and computationally expensive analysis, we are unable to derive these values for all publishers. We use 400 randomly selected publishers to represent the normal behaviour of all publishers and refer to this group as "All" in our analysis.

In order to compute these metrics, we need to estimate the time that a specific publisher has been connected to a torrent (in one or multiple sessions). Since each query to the tracker just reports (at most) a random subset of 200 IPs, in big torrents ($>$200 peers), we need to perform multiple queries in order to assess the presence of the publisher in the torrent. In the Appendix, we detail the technique used to estimate the session time of a specific user in each torrent.

Average Seeding Time: We measure the duration of time that a publisher stays in a torrent since its birth to seed the content. In general, a publisher can leave the torrent once there is an adequate fraction of other seeds. Fig. 4(a) depicts the summary distribution of average seeding time across all publishers in each target group. This figure demonstrates the following points. First, the seeding time for fake publishers is significantly longer than publishers in other groups. Since these publishers do not provide the actual content, the initial fake publisher remains as the only seed in the session (i.e., other users do not help in seeding fake content) to keep the torrent alive. Second, Fig. 4(a) shows that top publishers typically seed a content for a few hours. However, the seeding time for top publishers from hosting providers is clearly longer than top publishers from commercial ISPs. This suggests that publishers at hosting providers are more concerned about the availability of their published content.

Average Number of Parallel Torrents: Fig. 4(b) depicts the summary distribution of the average number of torrents that a publisher seeds in parallel across publishers in each target group. This figure indicates that fake publishers seed many torrents in parallel. We have seen that fake publishers typically publish a large number of torrents and other users do not help them for seeding. Therefore, fake publishers need to seed all of their seeded torrents in parallel in order to keep them alive. The results for top publishers show that their typical number of seeded torrents in parallel is the same (around three torrents) regardless of their location. However, we expect that a regular publisher seed only one file at a time.

Aggregated Session Time: We have also quantified the availability of individual publishers by estimating the aggregated session time that each publisher is present in the system across all published torrents. Fig. 4(c) shows the distribution of this availability measure across publishers in each target group. As expected, fake publishers present the longest aggregated session time due to their obligation to continuously seed their content to keep them alive. If we focus on top publishers, they exhibit a typical aggregated session 10 times longer than regular publishers. Furthermore, top publishers at hosting services are clearly more available than those at commercial ISPs.

D. Summary

BitTorrent content publishers can be broadly divided into three groups as follows.

1) Altruistic users publish content while consuming content that is published by other users.
2) Fake publishers publish a significant number of files that are often video and software content from a few hosting providers. Due to the fake nature of their content, their torrents are unpopular, and they need to seed all torrents to keep them alive. These publishers appear to be associated with antipiracy agencies or malicious users. We validate this hypothesis in Section V.
3) Top publishers publish a large number of popular files (often copyrighted video) and remain for a long time in the associated torrents to ensure a proper seeding of their published content. These publishers are located at hosting facilities or commercial ISPs. Their behavior suggests that these publishers are interested in the visibility of the published content possibly to attract a large number of users. The (cost of) allocated resources by these publishers along with legal implications of publishing copyrighted material cannot be considered as altruistic. Therefore, the most conceivable incentive for these publishers appears to be financial profit. We examine this hypothesis in the rest of this paper.

V. Incentives of Major Publishers

In this section, we examine the incentive of different groups of publishers in more detail. First, we focus on fake publishers by examining the name and content of the files they published. We noticed that these publishers often publish files with catchy titles (e.g., recently released Hollywood movies), and in most cases the actual content was not available when we tried to download it.[3] The few files we were able to download were indeed fake content. Some of them included an antipiracy message, whereas some others led to malicious software (or malware). In the latter case, the content was a video that had a pointer to a specific software (e.g., http://flvdirect.com/) to be downloaded in order to reproduce the video. This software was indeed a malware.[4] These observations support our hypothesis that fake publishers are either antipiracy agents who publish fake versions of copyrighted content or malicious users who lead users to download a malware. These clues reveal the likely incentives of most fake publishers. In a separate study [20], we characterize these fake publishers in more detail and explore possible solutions to mitigate their impact on consumers.

Second, another group of major publishers allocates a significant amount of resources to publish non-fake and often copyrighted content. We believe that the behavior of these users is not altruistic. More specifically, our hypothesis is that these publishers leverage major BitTorrent portals as a venue to freely attract consumers to their Web sites. To verify this hypothesis, we conduct an investigation to gather the following information about each one of the *top* (i.e., top-100 non-fake) publishers:

- *promoting URL:* the URL that downloaders of a published content may encounter;
- *publisher's username*: any publicly available information about the username that a major publisher uses in the Pirate Bay portal;
- *business profile*: offered services (and choices) at the promoting URL.

Next, we describe our approach for collecting this information.

Promoting URL: We emulate the experience of a user by downloading a few randomly selected files published by each top publisher to determine whether and where they may encounter a promoting URL. We identified three places where publishers may embed a promoting URL: 1) the name of the downloaded file (e.g., user mois20 names his files as *filename-divxatope.com*, thus advertising the URL www.divxatope.com); 2) the textbox in the Web page associated with each published content; 3) the name of a text file that is distributed with the actual content and is displayed by the BitTorrent software when opening the .torrent file. Our investigation indicates that the second approach (using the textbox) is the most common technique among the publishers.

Publisher's Username: We browsed the Internet to learn more information about the username associated with each top publisher. First, the username is in some cases directly related to the URL (e.g., user UltraTorrents whose URL is www.ultratorrents.com). This exercise also reveals whether this username publishes on other major BitTorrent portals in addition to the Pirate Bay. Finally, posted information in various forums could reveal (among other things) the promoting Web site.

Business Services: We characterize the type of services offered at the promoting URL and ways in which the Web site may generate income (e.g., posting ads). We also capture the exchanged HTTP headers between a Web browser and the promoting URL to identify any established connection to third-party Web sites (e.g., redirection to ads Web sites or some third party aggregator) using the technique described in [19].

A. Classifying Publishers

Using the above methodology, we examined a few published torrents for each one of the top publishers as well as sample torrents for 100 randomly selected publishers that are not in the top-100, called *regular publishers*. On the one hand, we did not discover any interesting or unusual behavior in torrents published by regular publishers and thus conclude that they behave in an altruistic manner. On the other hand, a large fraction of seeded torrents by the top publishers systematically promotes one or more Web sites with financial incentives. Our examination revealed that these publishers often include a promoting URL in the textbox of the content Web page. We classify these top publishers into the following three groups based on their type of business (using the content of their promoting Web sites) and describe how they leverage BitTorrent portals to intercept and redirect users to their Web sites.

Private BitTorrent Portals: A subset of major publishers (25% of top) own their BitTorrent portals that are in some cases associated with private trackers [16]. These private trackers offer a better user experience in terms of download

[3] We tried to download these files a few weeks after the correspondent measurement study was performed.

[4] http://www.prevx.com/filenames/X2669713580830956212-X1/FLVDIRECT.EXE.html.

rate (compared to major open BitTorrent portals), but require clients to maintain a certain seeding ratio. More specifically, each participating BitTorrent client is required to seed content proportionally to the amount of data that it downloads across multiple torrents. To achieve this goal, users are required to register in the Web site and login before downloading the torrent files. The publishers in this class publish 18% of all the content, while they are responsible for 29% of the downloads. Two thirds of these publishers advertise the URL in the textbox at the content Web page. Furthermore, they appear to gain financial profit in three different ways: 1) posting advertisement in their Web sites; 2) seeking donations from visitors to continue their basic service; and 3) collecting a fee for VIP access that allows the client to download any content without requiring any kind of seeding ratio. These publishers typically inject video, audio, and software content into BitTorrent portals. Interestingly, a significant fraction of publishers in this class (40%) publish content in a specific language (Italian, Dutch, Spanish, or Swedish), and specifically 66% of this group are dedicated to publishing Spanish content. This finding is consistent with prior reports on the high level of copyright infringement in Spain [11].

Promoting Web Sites: Another class of top publishers (23% of top) appears to be promoting some URLs that are associated with hosting images Web sites (e.g., www.pixsor.com), forums or even religious groups (e.g., lightmiddleway.com). These publishers inject 8% of the content and are responsible of 11% of the downloads. Most publishers in this class advertise their URL using the textbox in the content Web page. Furthermore, most of these publishers (70%), specifically those that are running a hosting image Web site, publish only pornographic content. Inspection of the associated hosting image Web sites revealed that they store adult pictures. Therefore, by publishing pornographic content in major BitTorrent portals, they are targeting a particular demography of users who are likely to be interested in their Web sites. The income of the portals within this class is based on advertisement.

Altruistic Publisher: The remaining top publishers (52% of top) appear to be altruistic users since they do not seem to directly promote any URL. These publishers are responsible of 11.5% of the content and roughly the same fraction of downloads. Many of these users publish small music and e-book files that require a smaller amount of seeding resources. Furthermore, they typically include a very extensive description of the content and often ask other users to help with seeding the content. These evidences suggest that these publishers may have limited resources, and thus they need the help of others to sustain the distribution of their content.

In summary, roughly half of the top publishers advertise a Web portal in their published torrents. It appears that their intention is to attract a large number of users to their Web sites. The income of these portals come from ads and in the specific case of private BitTorrent portals also from donations and VIP fees. Overall, these profit-driven publishers generate 26% of the content and 40% of the downloads. Therefore, the removal of this small fraction of publishers may have a significant impact on the popularity of major BitTorrent portals. Finally, a fraction of publishers appears to be altruistic and responsible for a notable fraction of published content and downloads (11.5%). This suggests that there are some seemingly ordinary users who dedicate their resources to share content with a large number of peers in spite of the potential legal implications of such activity.

TABLE IV
LIFETIME AND AVERAGE PUBLISHING RATE FOR THE DIFFERENT CLASSES OF CONTENT PUBLISHERS: BITTORRENT PORTALS, PROMOTING WEB SITES, AND ALTRUISTIC PUBLISHERS. THE REPRESENTED VALUES ARE MIN/AVG/MAX PER CLASS

	Lifetime (days)	Avg. Publishing Rate (torrents per day)
private portals	63/466/1816	0.57/11.43/79.91
promoting web sites	50/459/1989	0.38/4.31/18.98
altruistic	10/376/1899	0.10/3.80/23.67

B. Longitudinal View of Major Publishers

So far, we focused on the contribution of major publishers only during our measurement intervals. Having identified the top publishers in our *pb10* dataset, we examine the longitudinal view of the contribution by major publishers since they appeared on the Pirate Bay portal. Toward this end, for each top publisher, we obtain the username page on the Pirate Bay portal that maintains the information about all the published content and its published time by the corresponding user until our measurement date (June 4, 2010).[5] Using this information for all top publishers, we capture their publishing pattern over time with the following parameters: 1) *publisher lifetime*, which represents the number of days between the first and the last appearance of the publisher in the Pirate Bay portal; 2) *average publishing rate* that indicates the average number of published content per day during their lifetime.

Table IV shows the minimum/average/maximum values of these metrics for the different classes of publishers: private portals, promoting Web sites, and altruistic publishers. The profit-driven publishers (i.e., private portals and promoting Web sites) have been publishing content for 15 months on average (at the time of the measurement), while the most longed-lived ones have been feeding content for more than 5 years. Furthermore, some of these publishers exhibit a surprisingly high average rate of publishing content (80 files per day). The altruistic publishers present a shorter lifetime and a lower publishing rate that seems to be due to their weaker incentives and their limited availability of resources.

In summary, the lifetime of major publishers suggests that content publishing in BitTorrent seems to have been a profitable business for (at least) a couple of years. Furthermore, the high seeding activity by profit-driven publishers (e.g., private portals) over a long period of time implies a high and continuous investment for required resources that should be compensated by different types of income (e.g., ads) for these portals. We examine the income of the profit-driven publishers in Section V-C.

C. Estimating Publishers' Income

The evidence that we presented in previous sections suggests that the goal of half of the top publishers is to attract users to

[5]Note that we cannot collect this information about fake publishers since the Web pages of their associated publishers are removed by the Pirate Bay as soon as they are identified.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE V
PUBLISHER'S WEB SITE VALUE ($), DAILY INCOME ($), AND NUMBER OF DAILY VISITS FOR THE DIFFERENT CLASSES OF PROFIT-DRIVEN CONTENT PUBLISHERS: BITTORRENT PORTALS AND PROMOTING WEB SITES. THE REPRESENTED VALUES ARE MIN/MEDIAN/AVG/MAX PER CLASS

	Web site Value ($)	Web site Daily Income ($)	Web site Daily visits
private portals	1K/33K/313K/2.8M	1/55/440/3.7K	74/21K/174K/1.4M
promoting web sites	24/22K/142K/1.8M	1/51/205/1.9K	7/22K/73.5K/772K

their own Web sites. We also showed that most of these publishers promote conditions to generate income by posting ads in their Web sites. In essence, these publishers have a clear financial incentive to publish content. In order to validate this key point, we assess their ability to generate income by estimating three important but related properties of their promoting Web sites: 1) average value of the Web site; 2) average daily income of the Web site; and 3) average daily visits to the Web site.

In the absence of a reliable source to obtain this sensitive information, we rely on several companies[6] that monitor and report these statistics for different Web sites. Since these companies do not reveal the details of their monitoring strategy, we cannot assess the accuracy of their reported statistics. To reduce any potential error in the provided statistics by individual companies, we collect this information from six independent companies and use the average value of these statistics among them. We emphasize that the obtained statistics from these companies are treated as ballpark estimates for the three properties of the promoting Web sites to enable our validation.

Table V presents the minimum, median, average, and maximum values of the described metrics for each class of profit-driven publisher classes (i.e., private portals and promoting Web sites). The median values suggest that the promoted Web sites are fairly profitable since they value tens of thousands of dollars with daily income of a few hundred dollars and tens of thousands of visits per day. Furthermore, few publishers (<10) are associated to very profitable Web sites valued in hundreds of thousands to millions of dollars, which receive daily income of thousands of dollars and hundreds of thousands of visits per day.

In summary, these statistics suggest that these Web sites are valuable and visible and generate a substantial level of income.

VI. EXAMINING CONSUMER LOYALTY

In this section, we examine the behavior of consumers toward individual top publishers and their relationship with publishers' profile. More specifically, our goal is to answer two basic questions.

- Can individual top publishers attract loyal consumers that primarily download content from that top publisher?
- Is there any correlation between a publisher's profile and the level of loyalty among its consumers?

In essence, answering these questions reveals whether top publishers adopt business practices that affect their ability to form a loyal consumer base within the BitTorrent ecosystem in order to achieve their goals. The notation used in this section is shown in Table VI.

[6]sitelogr.com, cwire.com, websiteoutlook.com, sitevaluecalculator.com, mywebsiteworth.com, yourwebsitevalue.com.

TABLE VI
NOTATION USED IN SECTION VI (EXAMINING CONSUMER LOYALTY)

Metric	Definition
pub(c)	Num. of publishers from which consumer c downloads files
dl(c)	Num. of files downloaded by consumer c
C(c)	Num. of files downloaded by consumer c from its preferred publisher
L(c)	Normalized loyalty of consumer c to its preferred publisher
NLC(p)	Absolute number of loyal consumers for publisher p
FLC(p)	Fraction of loyal consumers for publisher p

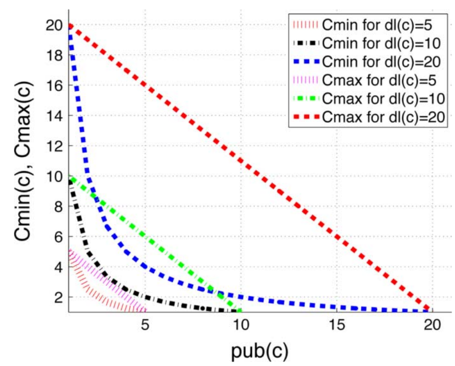


Fig. 5. Max and min values for $C(c)$ for different $\mathrm{pub}(c)$ and $\mathrm{dl}(c)$ values.

A. Quantifying Consumer Loyalty

To study the loyalty of consumers, we need to define a meaningful metric to quantify this attribute of a consumer toward a particular publisher. Suppose that consumer c downloads $\mathrm{dl}(c)$ files from $\mathrm{pub}(c)$ different publishers where the largest number of downloaded files by c from a publisher is $C(c)$ ($C(c) \leq \mathrm{dl}(c)$). Consumer c is considered loyal to publisher p if it downloads most of its files from p. We refer to p as the *preferred* publisher for consumer c.

On the one hand, if the downloaded files by consumer c are evenly divided among $\mathrm{pub}(c)$ publishers, c is *not* loyal to any publisher since it shows the minimum consumption level ($C_{\min}(c)$) toward any publisher that is simply

$$C_{\min}(c) = \frac{\mathrm{dl}(c)}{\mathrm{pub}(c)}. \quad (1)$$

On the other hand, for a given consumer c, the consumption from its preferred publisher is maximized when c downloads only one file from each nonpreferred publisher and all remaining files from its preferred publisher. Thus $C_{\max}(c)$ can be easily expressed as

$$C_{\max}(c) = \mathrm{dl}(c) - \mathrm{pub}(c) + 1. \quad (2)$$

Given a particular scenario defined by $\mathrm{dl}(c)$ and $\mathrm{pub}(c)$, the above simple equations for $C_{\min}(c)$ and $C_{\max}(c)$ determine the possible range for the number of files that a user c can download from a publisher. Fig. 5 shows the variation of $C_{\min}(c)$ and $C_{\max}(c)$ as a function of $\mathrm{pub}(c)$ for different $\mathrm{dl}(c)$. This figure reveals that both $C_{\min}(c)$ and $C_{\max}(c)$ can significantly change across different scenarios as $\mathrm{dl}(c)$ and $\mathrm{pub}(c)$ vary. To allow comparison of the loyalty of users across different scenarios,

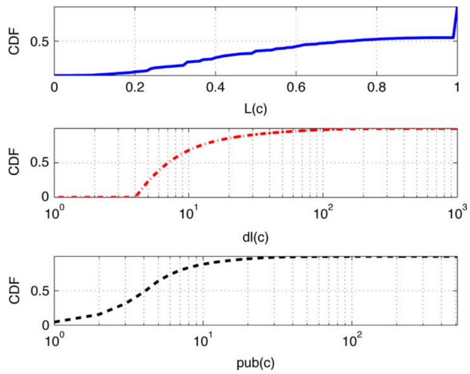


Fig. 6. Distribution of $\mathrm{L}(c)$, $\mathrm{dl}(c)$, and $\mathrm{pub}(c)$ across all consumers with $\mathrm{dl}(c) \geq 5$.

we define the loyalty of user c toward his preferred publisher as his normalized consumption as follows:

$$\mathrm{L}(c) = \frac{C(c) - C_{\min}(c)}{C_{\max}(c) - C_{\min}(c)} \qquad \text{so that } 0 \leq \mathrm{L}(c) \leq 1. \quad (3)$$

We use our *pb10* dataset for this analysis. This dataset contains 27.3 M consumers, however we only focus on 2.6 M consumers that are moderately active (i.e., have downloaded at least five files, $\mathrm{dl}(c) \geq 5$) and exhibit a positive loyalty ($\mathrm{L}(c){>}0$). Fig. 6 shows the distribution of $\mathrm{pub}(c)$, $\mathrm{dl}(c)$, and $\mathrm{L}(c)$ among these consumers. On the one hand, we observe that a majority (85%) of these consumers download less than 19 files and from less than eight different publishers during our one-month measurement period. This figure shows that roughly 45% of these consumers exhibit a $\mathrm{L}(c)$ value very close to 1, and the median $\mathrm{L}(c)$ value is 0.73. This suggests that half of these consumers exhibit a rather high level of loyalty toward a particular publisher. For the rest of the analysis in this section, we focus on all moderately active consumers with positive loyalty and their corresponding publishers. We also filter consumers based on their $\mathrm{dl}(c)$ values to examine more active consumers.

B. Consumer Loyalty Among Publishers

We first examine the level of loyalty among consumers of two group of publishers (as our target groups) as follows:

1) *top-100*: the non-fake top-100 publishers that we identified in Section III. We recall that only 84 of the top-100 publishers were non-fake;
2) *random*: 100 non-fake randomly selected publishers (excluding the top-100) to represent the rest of publishers in our dataset.

We define loyalty for individual consumers. We need to introduce two metrics to assess different aspects of loyalty of consumers toward a particular publisher p as follows:

- $\mathrm{NLC}(p)$: the absolute number of loyal consumers for p;
- $\mathrm{FLC}(p)$: the fraction of p's consumers that are loyal.

For publisher p, $\mathrm{FLC}(p)$ indicates what fraction of p's consumers are loyal to p, while $\mathrm{NLC}(p)$ measures how many loyal consumers p has. To clarify the relation between these two metrics, let us consider the following simple example: Publisher p_1 with 1000 consumers and $\mathrm{NLC}(p_1) = 100$ has $\mathrm{FLC}(p_1) = 0.1$, whereas publisher p_2 with 200 consumers and $\mathrm{NLC}(p_2) = 100$

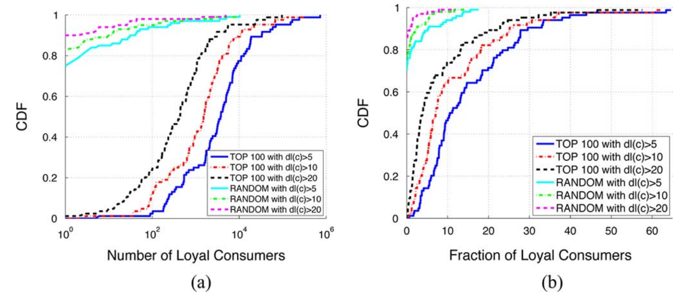


Fig. 7. Cumulative distribution function (CDF) of $\mathrm{NLC}(p)$ and $\mathrm{FLC}(p)$ for non-fake top-100 publishers. (a) NLC. (b) FLC.

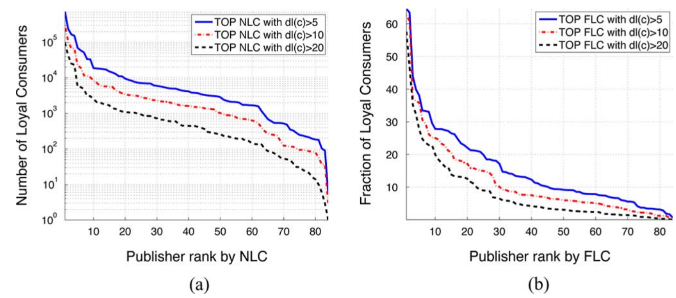


Fig. 8. $\mathrm{NLC}(p)$ and $\mathrm{FLC}(p)$ for non-fake top-100 publishers. x-axis indicates the publisher's rank based on the target metric among publishers. (a) NLC. (b) FLC.

has $\mathrm{FLC}(p_2) = 0.5$. Both p_1 and p_2 are able to attract the same number of loyal consumers, however the *strategy* used by p_2 seems to be more effective since a higher percentage of all its consumers is loyal.

Fig. 7(a) and (b) depicts the distribution of $\mathrm{NLC}(p)$ and $\mathrm{FLC}(p)$ across all publishers in each one of the target groups, respectively. Furthermore, each figure also plots the specified distribution by considering only a subset of consumers whose $\mathrm{dl}(c)$ is larger than 5, 10, and 20.

These figures demonstrate that the top-100 publishers not only have a significantly larger number of loyal consumers, but also have a larger fraction of loyal consumers. Note that as we focus on more active consumers (with larger $\mathrm{dl}(c)$), the distributions of $\mathrm{NLC}(p)$ and $\mathrm{FLC}(p)$ maintain the same shape but shift toward lower values for both groups of publishers. This suggests that the observed trends among target groups of publishers do not significantly change by considering consumers with different level of activities.

For the rest of this section, we focus on the top-100 publishers since a majority of all loyal consumers are associated with these publishers. In particular, 72%, 64%, and 48% of consumers are associated with these publishers for $\mathrm{dl}(c)$ larger or equal to 5, 10, and 20, respectively. In our analysis, we will also leverage the business profile of these publishers that we determined in Section V.

C. Loyalty Toward Top-100 Publishers

Focusing our analysis on the top-100 publishers, Fig. 8(a) and (b) shows the value of $\mathrm{NLC}(p)$ and $\mathrm{FLC}(p)$ across these publishers, respectively. In these figures, publishers across the x-axis are ranked by their $\mathrm{NLC}(p)$ (or $\mathrm{FLC}(p)$) value, and each line shows the result using a different minimum $\mathrm{dl}(c)$ value for

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

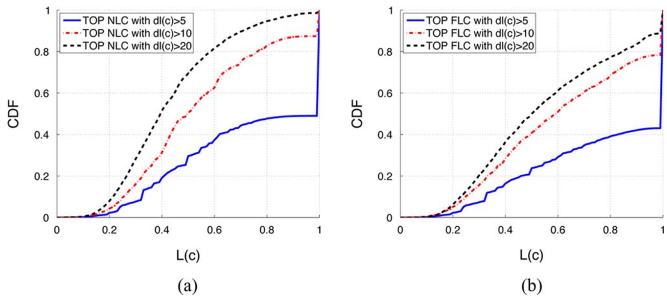


Fig. 9. CDF of $\mathrm{L}(c)$ across consumers of top-NLC and top-FLC publishers using different minimum $\mathrm{dl}(c)$ to filter consumers. (a) top-NLC. (b) top-FLC.

filtering active consumers. Note that the y-axis in Fig. 8(a) has log scale.

These figures reveal that the values of $\mathrm{NLC}(p)$ and $\mathrm{FLC}(p)$ across the top-100 publishers vary dramatically. In particular, the top-10 publishers with the largest $\mathrm{NLC}(p)$ values collectively attract around 84% of all loyal consumers associated with all top-100 publishers (independent of the minimum level of activity among consumers). We call this group *top-NLC*. Furthermore, only top-10 publishers based on $\mathrm{FLC}(p)$ have a significant fraction of loyal consumers (at least 14%–28% for different $\mathrm{dl}(c)$). We call this group *top-FLC*. Focusing on more active consumers obviously reduces the number of loyal consumers and thus the value of $\mathrm{NLC}(p)$ and $\mathrm{FLC}(p)$ for each publisher. However, increasing $\mathrm{dl}(c)$ does not seem to generally change the overall trends of these results. This suggests that only top-NLC and top-FLC publishers appear to have a significant base of loyal consumers, and thus we focus on these two groups.[7]

Examinations of the identity of top-NLC and top-FLC publishers for different $\mathrm{dl}(c)$ values reveals the following key points. First, there are only two overlapping publishers, namely *eztv* and *exmnova*, between two groups for $\mathrm{dl}(c) \geq 5$. Second, as we focus on more active consumers $(\mathrm{dl}(c) \geq 10, 20)$, we observe only two other overlapping publishers, namely *Rabiner* and *artpepper*, in top-NLC and top-FLC groups, but they are ranked at the end of these groups.

In summary, our results show that consumer loyalty toward publishers (measured by $\mathrm{NLC}(p)$ *or* $\mathrm{FLC}(p)$*) is very skewed regardless of the minimum expected level of activity among consumers. Only a small number of top-NLC and top-FLC publishers appears to have a significant base of loyal consumers. However, most of the publishers in these two groups are unique. This suggests that top-NLC and top-FLC publishers are likely to exhibit different characteristics*. Therefore, we investigate our two motivating questions for nonoverlapping top-NLC and top-FLC publishers and their consumers.

Fine-Grain Loyalty Toward Top Publishers: So far, we only considered $\mathrm{NLC}(p)$ and $\mathrm{FLC}(p)$ as two coarse measures of consumer loyalty toward publisher p. We now take a closer look at the level of loyalty by individual consumers (or $\mathrm{L}(c)$) toward top-NLC and top-FLC publishers. Fig. 9(a) and (b) depicts the distribution of $\mathrm{L}(c)$ among all loyal consumers of nonoverlapping top-NLC and top-FLC publishers, respectively. Each figure shows the distribution for different minimum level of activity $(\mathrm{dl}(c))$ among consumers as well. Comparing lines for similar $\mathrm{dl}(c)$ values in these figures demonstrates that *the top-FLC publishers not only attract a higher fraction of loyal consumers, but the level of loyalty among their consumers is relatively higher than in the case of top-NLC publishers*.

[7]We have also identified and examined the top-10 publishers based on FLC across *all* publishers (not just top-100). While two of these publishers are in our top-FLC, the other eight publish a very small number of files and attract an insignificant number of consumers. Since their impact is negligible, we only focus on top-FLC.

TABLE VII
AVERAGE LOYALTY FOR CONSUMERS OF TOP-FLC AND TOP-NLC PUBLISHERS

dl(c)	Avg(L(c)) Top-NLC	Avg(L(c)) Top-FLC	Norm Diff
≥ 5	0.72	0.74	2.74%
≥ 10	0.54	0.61	12.17%
≥ 20	0.40	0.56	33.33%

To better demonstrate this point, we use the average value of $\mathrm{L}(c)$ across consumers of publishers in each group $(\mathrm{Avg}(\mathrm{L}(c)))$. Since the value of $\mathrm{L}(c)$ is always between 0 and 1, average $\mathrm{L}(c)$ provides a useful indicator to compare two groups. Table VII summarizes the average value of $\mathrm{L}(c)$ across consumers for top-NLC and top-FLC publishers using different $\mathrm{dl}(c)$ values for filtering. The last column of Table VII presents the normalized difference in average $\mathrm{L}(c)$ between these two groups. This table clearly shows the following points: 1) The average loyalty among consumers of top-FLC consumers is higher than for consumers of top-NLC for any minimum level of activity among consumers; 2) the value of $\mathrm{NormDiff}$ reveals that the gap between loyalty of consumers grows as we focus on more active consumers.

D. *Top-NLC Versus Top-FLC Publishers*

Our hypothesis is that top-FLC and top-NLC publishers exhibit different business profiles, which in turn results in their high $\mathrm{FLC}(p)$ or $\mathrm{NLC}(p)$ values. To explore this hypothesis, we examined the following key attributes of these two groups of publishers: their username, business profile (as we determined in Section V), type of posted content, assessed reputation by the Pirate Bay portal, number of published content, and the total number of their consumers. In particular, the level of reputation for each publisher is assigned by the Pirate Bay based on the past history of the publisher on this portal. These levels from high to low are as follows: VIP, Trusted, Helper, and Unknown (i.e., no reputation is assigned). Since the information about individual publishers was collected a few months after our main data collection, some of the publishers have left the Pirate Bay. The reputation of these departed publishers was set to *Deleted*. Tables VIII and IX provide detailed characteristics of top-NLC and top-FLC publishers. For easier comparison, Table X summarizes the range of these characteristics (with the median value in parentheses) for nonoverlapping publishers in both groups.

Interestingly, uncommon publishers in each group exhibit rather distinct characteristics. Nonoverlapping top-NLC publishers are mostly profit-driven publishers that publish 47–2332 popular content (e.g., recent episodes of popular TV shows, recent Hollywood movies, or porn videos). In addition, the ranges of $\mathrm{FLC}(p)$ and $\mathrm{NLC}(p)$ values for these publishers are 9%–27% and 18 K–166 K, respectively. In contrast, nonoverlapping top-FLC publishers are mostly altruistic publishers

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE VIII
MAIN CHARACTERISTICS OF TOP-NLC PUBLISHERS FOR $\mathrm{dl}(c) \geq 5$; PP: PRIVATE (BITTORRENT) PORTAL, PROMO: PROMOTING WEB SITE, A: ALTRUISTIC

USERNAME	BUSINESS	TYPE OF PUBLISHED CONTENT	REPUTATION AT PIRATE BAY	NUMBER OF PUBLISHED CONTENT	NUMBER OF CONSUMERS	NLC(p)	FLC(p)	AVG(L(c))
eztv	PP	VIDEO (Tv Shows)	VIP	313	1,151,633	730,558	63.44	0.7532
exmnova	PP	PORN (Movies)	Trusted	1,780	633,995	241,003	38.01	0.6677
TvTeam	PP	VIDEO (Tv Shows)	VIP	2,332	799,035	166,497	20.84	0.7664
Rabiner	PROMO	PORN (Movies)	Trusted	662	559,526	152,342	27.23	0.7187
raymondhome	PROMO	VIDEO (Movies)	VIP	125	494,343	69,451	14.05	0.7816
VTV	A	VIDEO (Tv Shows)	VIP	119	632,672	59,029	9.33	0.7144
extremezone	PP	VIDEO (Movies)	VIP	47	221,430	51,832	23.41	0.8435
Housezz	A	AUDIO, APPS, VIDEO	Deleted	225	332,959	33,609	10.09	0.7054
pizstol	PROMO	PORN (Movies)	Deleted	461	292,708	33,191	11.34	0.6669
1.No.1	PROMO	PORN (Movies)	VIP	223	200,857	18,812	9.37	0.5563

TABLE IX
MAIN CHARACTERISTICS OF TOP-FLC PUBLISHERS FOR $\mathrm{dl}(c) \geq 5$; PP: PRIVATE (BITTORRENT) PORTAL, PROMO: PROMOTING WEB SITE, A: ALTRUISTIC

USERNAME	BUSINESS	TYPE OF PUBLISHED CONTENT	REPUTATION AT PIRATE BAY	NUMBER OF PUBLISHED CONTENT	NUMBER OF CONSUMERS	NLC(p)	FLC(p)	AVG(L(c))
ClaudiaZ	A	VIDEO (Tv Shows)	VIP	162	7,730	4,994	64.61	0.7882
eztv	PP	VIDEO (Tv Shows)	VIP	313	1,151,633	730,558	63.44	0.7532
Mois20	PP	VIDEO (Tv Shows)	VIP	250	38,919	17,099	43.94	0.6677
CanadaJoe	A	AUDIO	VIP	401	14,458	5,768	39.90	0.7853
exmnova	PP	PORN (Movies)	Trusted	1,780	633,995	241,003	38.01	0.6677
mikexxxryan	A	PORN (Movies)	Deleted	61	17,796	5,952	33.45	0.8645
starburst3	A	PORN (Pictures)	Deleted	133	7,315	2,436	33.30	0.6568
0oEdito0	A	COMICS	Deleted	73	18,616	6,153	33.05	0.5780
SkullManWoopt	A	PORN (Picture)	Trusted	69	12,502	3,702	29.61	0.7527
ESPALPSP	A	GAMES	Deleted	90	31,728	8,829	27.83	0.8081

TABLE X
MAIN CHARACTERISTICS OF TOP-FLC AND TOP-NLC PUBLISHERS

Attribute	top-FLC	top-NLC
$FLC(p)$	27-65% (33%)	9-27% (13%)
$NLC(p)$	3k-17k (6k)	18k-166k (55k)
# Pub. Files	61-401 (111)	47-2332 (224)
# Consumers	7k-39k (16k)	200k-799k (413k)
Profile Type	Altruistic	Private Tracker
Content Type	Foreign, Comic	Movie/TV Show

who upload a small to moderate number (61–401) of rather specialized content (e.g., movies in a non-English language, comics, or porn pictures). Their published content is not as popular as for top-NLC publishers, thus they attract a smaller number of consumers and therefore a smaller number of loyal consumers (3 K–17 K). However, a larger fraction $(\mathrm{FLC}(p))$ of their consumers are loyal and exhibit a rather larger level of loyalty than consumers of top-NLC. The overlapping publishers appear to exhibit a combination of these characteristics, which results in their appearance in both groups.

In summary, top-FLC and top-NLC publishers exhibit different characteristics in terms of the number, type, and popularity of published content that lead to a different loyalty pattern among their consumers.

VII. OTHER BENEFICIARIES IN THE BITTORRENT MARKETPLACE

In previous sections, we analyzed the main characteristics of major content publishers in large BitTorrent portals, demonstrating that content publishing appears to be a profitable *business* for an important fraction of the top publishers. While we have focused primarily on content publishers, there are other players around the BitTorrent ecosystem [28] that have financial interest and may promote this marketplace around BitTorrent. These other beneficiaries include: *major BitTorrent portals*, *hosting providers*, and *ad companies*. Fig. 10 shows the interactions between different players in the BitTorrent marketplace, where the arrows indicate the flow of money between them. In this section, we briefly describe the role of main players and their interactions with others.

Major public BitTorrent portals such as the Pirate Bay are dedicated to indexing torrent files. They basically serve as rendezvous points for content publishers and consumers. The main advantage of these major portals is the reliable access (e.g., by rapidly removing fake or infected content) to popular content. This motivates millions of BitTorrent users to visit these portals every day, which in turn makes these Web sites very valuable. For instance, the Pirate Bay is one of the most popular sites across Internet (ranked the 77th in the Alexa Ranking as of November 15, 2011) and is valued around $10 M.

Hosting providers are companies dedicated to renting servers. The heavy seeding activity performed by some publishers requires significant resources (e.g., bandwidth and storage). Thus, a large fraction of major publishers rent servers from hosting providers that generates income for hosting providers proportional to the level of activity by the publisher. For example, our measurement revealed that around 78–164 servers (i.e., unique IP addresses) associated with major publishers are hosted at a single ISP in France, called OVH. Considering the cost of an average server at OVH (around 300€/month from OVH Web site), we estimate that its average monthly income from BitTorrent publishers is between 23 K to 43 K€/month. It is worth noting that some hosting providers (e.g., Server Intellect) have adopted strict policies against P2P applications using their servers to

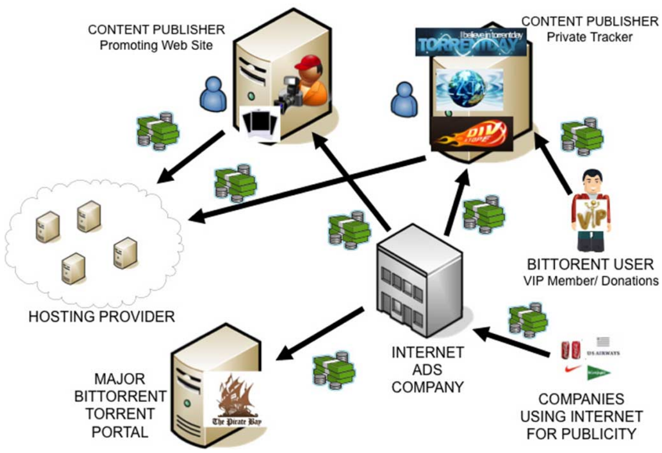


Fig. 10. Business model of content publishing in Bittorrent.

distribute copyrighted material due to possible legal implications [9]. However, our exchange with OVH revealed that they do not monitor the activities performed by their customers and may react only when a violation is reported by a third party and if the related activity is not ceased by the customer [7]. This reactive and rather soft policy appears to have attracted publishers of copyrighted content to OVH.

Ad companies pair customers who wish to post ads on the Internet with popular Web sites where ads can be placed. These companies dynamically determine to which, typically popular, Web site and when each ad is placed. The ad company and the Web site both receive a portion of this income. By attracting users through major BitTorrent portals, content publishers can increase the number of visits to their Web sites and thus become a more desirable target for posting ads. We have examined the header of exchanged http messages between the browser and the publishers' Web sites and verified that these Web sites indeed host ads. Unfortunately, we are unable to estimate the level of income that publishers have from hosting ads.

VIII. Software for Monitoring Content Publishing

In order to make our measurement techniques and our findings more accessible to other researchers and BitTorrent users, we have integrated our measurement tools into a system called *Monitoring, identifYing & PROfiling BitTorrent publishErs* (MYPROBE). MYPROBE continuously monitors the publishing activity in the Pirate Bay portal by implementing the measurement methodology that we described in Section II. In particular, it leverages the RSS feed to quickly detect a newly published content and then retrieves the following information for a detected torrent: filename, content category and subcategory (based on the Pirate Bay categories), publisher's username, and (in those cases that we can) the publisher's IP address as well as the ISP, city, and country associated to this IP address. The main characteristics of any identified torrent and its publisher are stored in a database with a Web-based portal (at [3]) that allows users to query and obtain information. Furthermore, access to the dataset used in this paper is granted to interested researchers by contacting the authors. Finally, it is worth adding that we have also designed and implemented a mechanism that enables BitTorrent users to identify and avoid fake publishers [20].

IX. Related Work

A significant research effort has focused on understanding different aspects of BitTorrent by gathering data from live swarms. Most of these studies have primarily examined demographics of users [17], [26], [28] and technical aspects of swarming mechanisms [14], [18], [24]. However, to our knowledge, the socioeconomic aspects of BitTorrent that we addressed in this paper have received little attention. The most relevant work to this paper is a recent study that examined the weakness of BitTorrent privacy [12]. The authors analyzed the demography of BitTorrent content publishers and presented a highly skewed distribution of published content among them as well as the presence of a significant fraction of publishers located at hosting providers. This indeed validates some of our initial observations. In another study, Zhang *et al.* [28] presented the most extensive characterization of the BitTorrent ecosystem. This study briefly examined the demography of content publishers and showed a skewed distribution of the contributed content among them. The authors identify the publishers by their usernames. We have shown that this assumption may miss an important group of publishers who post fake content, i.e., fake publishers. Our work goes beyond the simple examination of demographics of content publishers. We identify, characterize, and classify the major publishers and more interestingly reveal their incentives and their motivating business model.

X. Conclusion

In this paper, we present a measurement study on the largest BitTorrent portals to investigate socioeconomic incentives among content publishers. Our results revealed that a small fraction of publishers is responsible for two thirds of the published content and three quarters of the downloads. Our careful investigation on the incentives of major publishers in the largest BitTorrent portals led to the following important findings. First, anti-piracy agencies and malicious users perform a systematic poisoning attack over major BitTorrent portals by publishing fake content in order to obstruct download of copyrighted content and to spread malware, respectively. Roughly one third of the published content and a quarter of all downloads are associated with fake content. This finding indicates that BitTorrent portals can be leveraged by malicious users to easily spread malware to a large number of users, which could be a major security concern. Second, excluding the fake publishers, the remaining top publishers are responsible for one third of all published content and half of all the downloads. Our evidences suggest that half of these top publishers leverage the published copyrighted content on BitTorrent portals to attract content consumers to their Web sites for financial gain. We also demonstrate that these profit-driven publishers exhibit clearly distinct characteristics (i.e., a signature) compared to altruistic publishers. Third, we examined consumer loyalty toward top publishers and showed that the altruistic publishers attract a larger fraction of loyal consumers with a higher level of loyalty compare to profit-driven publishers.

Overall, our study sheds an insightful light on socioeconomic factors that seem to drive the popularity of BitTorrent and thus could affect the significant impact of its associated traffic on the Internet.

Appendix
Estimation of Session Duration

In this Appendix, we explain the procedure utilized to calculate the duration of the session time of a given peer in a given torrent. We explain the procedure using our *mn08* dataset. Note that it would be similar for *pb10*.

Our *mn08* crawler connects to the tracker periodically and obtains a random subset of all the IP addresses participating in the torrent. Then, we cannot guarantee to obtain the IP address of the target peer in a resolution of seconds or even few minutes. This imposes some restrictions to compute the content publisher's seeding time in a given torrent.

Therefore, we first define a model to estimate the number of queries to the tracker (m) needed to obtain the IP address of the content publisher with a given probability $\mathcal{P}$. Let us assume that: 1) we have a torrent with $\mathcal{N}$ peers; and 2) for each query, the tracker gives us a random set of $\mathcal{W}$ IP addresses. Then, if the target peer is in the torrent, the probability $(\mathcal{P})$ of obtaining its IP address in m consecutive queries to the tracker is given by

$$\mathcal{P} = 1 - \left(1 - \frac{\mathcal{W}}{\mathcal{N}}\right)^m. \quad (4)$$

We have computed the maximum instantaneous population of the torrents in our *mn08* and found that 90% of the torrents have typically less than 165 concurrent peers. Then, we assume that the torrents have always a population $\mathcal{N} = 165$. This is an upper bound that allows us to remove the noise introduced by the churn. We make a second conservative assumption: The tracker gives us $\mathcal{W} = 50$ random IPs in each response (although in most of the cases we obtain 200 IP addresses). With these numbers and the proposed model, we can assure that if a peer (e.g., a content publisher) is in the torrent, we will discover it in $m = 13$ queries to the tracker with a probability higher than 0.99.

Next, we have calculated the time between two consecutive queries to the tracker in our dataset and have checked that 90% of them are less than 18 min apart. Then, we again make a conservative assumption and consider that the time between two consecutive queries is 18 min.

Hence, multiplying the number of needed queries by the time between two consecutive queries, we conclude that if a peer (e.g., content publisher) is in the torrent, we are able to get its IP address in a period of 4 h with a probability equal to 0.99. Therefore, we consider that a given content publisher is offline (i.e., its session has finished) if its IP address is not gathered in the torrent during 4 h. We have repeated the experiments with 2- and 6-h thresholds, obtaining similar results.

Acknowledgment

The authors would like to thank anonymous International Conference on Emerging Networking Experiments and Technologies (CoNEXT) and IEEE/ACM Transactions on Networking reviewers for their valuable feedback, as well as J. Crowcroft for his insightful comments. They are also grateful to D. Rubio and A. González for help on developing the Web-based application.

References

[1] C. Williams, "UK cops arrest six alleged BitTorrent music uploaders," 2008 [Online]. Available: http://www.theregister.co.uk/2008/06/02/onk_further_arrests/
[2] "First BitTorrent arrest in Hong Kong," 2005 [Online]. Available: http://yro.slashdot.org/yro/05/01/13/2248252.shtml?id=123&tid=97&tid=95&tid=1
[3] "MYPROBE," [Online]. Available: http://btpublishers.netcom.it.uc3m.es
[4] Alexa, San Francisco, CA, "Alexa ranking," [Online]. Available: http://www.alexa.com/topsites
[5] MaxMind, Waltham, MA, "MaxMind—GeoIP," [Online]. Available: http://www.maxmind.com/app/ip-location
[6] Mininova, Utrecht, The Netherlands, "Mininova," [Online]. Available: http://mininova.org
[7] OVH, Ltd., London, U.K., "OVH's terms of service," [Online]. Available: http://www.ovh.co.uk/support/termsofservice/
[8] Sandvine, Waterloo, Canada, "Sandvine Fall 2011 global Internet phenomena report," 2011 [Online]. Available: http://www.sandvine.com/news/global_broadband_trends.asp
[9] Server Intellect, Orlando, FL, "Server Intellect use policy," [Online]. Available: http://www.serverintellect.com/terms/aup.aspx
[10] "The Pirate Bay," [Online]. Available: http://thepiratebay.se
[11] BayTSP, Los Gatos, CA, "BayTSP annual report," 2008 [Online]. Available: http://tech.mit.edu/V129/N28/piracy/BayTSP2008report.pdf
[12] S. Le Blond, A. Legout, F. Lefessant, W. Dabbous, and M. Ali Kaafar, "Spying the world from your laptop," in *Proc. USENIX LEET*, 2010.
[13] D. R. Choffnes, J. Duch, D. Malmgren, R. Guimera, F. E. Bustamante, and L. Amaral, "Strange bedfellows: Communities in BitTorrent," in *Proc. IPTPS*, 2010, Article no. 13.
[14] R. Cuevas, N. Laoutaris, X. Yang, G. Siganos, and P. Rodriguez, "Deep diving into BitTorrent locality," in *Proc. IEEE INFOCOM*, 2011, pp. 963–971.
[15] L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, and X. Zhang, "Measurements, analysis, and modeling of BitTorrent-like systems," in *Proc. ACM IMC*, 2005, p. 4.
[16] D. Hales, R. Rahman, B. Zhang, M. Meulpolder, and J. Pouwelse, "BitTorrent or BitCrunch: Evidence of a credit squeeze in BitTorrent?," in *Proc. IEEE WETICE*, 2009, pp. 99–104.
[17] M. Izal, G. Urvoy-Keller, E. W. Biersack, P. A. Felber, A. A. Hamra, and L. Garces-Erice, "Dissecting bittorrent: Five months in a torrent's lifetime," in *Proc. PAM*, 2004, pp. 1–11.
[18] S. Kaune, R. Cuevas, G. Tyson, A. Mauthe, C. Guerrero, and R. Steinmetz, "Unraveling BitTorrent's file unavailability: Measurements and analysis," in *Proc. IEEE P2P*, 2010, pp. 1–9.
[19] B. Krishnamurthy and C. E. Wills, "On the leakage of personally identifiable information via online social networks," in *Proc. ACM WOSN*, 2009, pp. 7–12.
[20] M. Kryczka, R. Cuevas, R. Gonzalez, A. Cuevas, and A. Azcorra, "TorrentGuard: Stopping scam and malware distribution in the BitTorrent ecosystem," Tech. rep., 2011 [Online]. Available: http://arxiv.org/abs/1105.3671
[21] N. Laoutaris, D. Carra, and P. Michiardi, "Uplink allocation beyond choke/unchoke or how to divide and conquer best," in *Proc. ACM CoNEXT*, 2008, Article no. 18.
[22] A. Legout, N. Liogkas, E. Kohler, and L. Zhang, "Clustering and sharing incentives in BitTorrent systems," in *Proc. ACM SIGMETRICS*, 2007, pp. 301–312.
[23] J. Liang, N. Naoumov, and K. W. Ross, "The index poisoning attack in P2P file sharing systems," in *Proc. IEEE INFOCOM*, 2006, pp. 1–12.
[24] D. Menasche, A. Rocha, B. Li, D. Towsley, and A. Venkataramani, "Content availability in swarming systems: Models, measurements and bundling implications," in *Proc. ACM CoNEXT*, 2009, pp. 121–132.
[25] M. Piatek, T. Isdal, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "Do incentives build robustness in BitTorrent?," in *Proc. USENIX NSDI*, 2007, pp. 1–14.
[26] J. A. Pouwelse, P. Garbacki, D. H. J. Epema, and H. J. Sips, "The BitTorrent P2P file-sharing system: Measurements and analysis," in *Proc. IPTPS*, 2005, pp. 205–216.

[27] M. Sirivianos, J. Han, P. Rex, and C. X. Yang, “Free-riding in BitTorrent networks with the large view exploit,” in *Proc. IPTPS*, 2007.

[28] C. Zhang, P. Dhungel, D. Wu, and K. W. Ross, “Unraveling the BitTorrent ecosystem,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 7, pp. 1164–1177, Jul. 2010.

Ruben Cuevas (M’12) received the M.Sc. and Ph.D. degrees in telematics engineering from the University Carlos III of Madrid, Leganes, Spain, in 2007 and 2010, respectively.

He is currently an Assistant Professor with the Telematics Engineering Department, University Carlos III of Madrid. He has been a Research Intern with Telefonica Research Lab, Barcelona, Spain, and a Courtesy Assistant Professor with the University of Oregon, Eugene, in 2008 and 2012, respectively. He has been involved in several national and international research projects in the area of content distribution and green networking. Moreover, he is coauthor of more than 30 papers in prestigious international journals and conferences such as the IEEE/ACM TRANSACTIONS ON NETWORKING, ACM CoNEXT, IEEE INFOCOM, and IEEE P2P. His main research interests include content distribution, P2P networks, online social networks, and Internet measurements.

Michal Kryczka received the Master’s degree in computer science from the Technical University of Lodz, Lodz, Poland, in 2008, and the Master’s degree in telematics engineering from the University Carlos III of Madrid, Leganes, Spain, in 2009, and is currently pursuing the Ph.D. degree in telematics engineering at the University Carlos III of Madrid.

Since 2008, he has been a Research Assistant (financed by an FPU scholarship from the Spanish Ministry of Education) with the Institute IMDEA Networks, Madrid, Spain. He is the coauthor of several papers in international conferences (ACM CoNEXT, IEEE P2P) and journals (IEEE/ACM TRANSACTIONS ON NETWORKING, IEEE COMMUNICATIONS LETTERS). His main research interests include P2P networks and Internet measurements.

Angel Cuevas (M’11) received the Master’s degree in telecommunication engineering and M.Sc. and Ph.D. degrees in telematics engineering from the University Carlos III of Madrid, Leganes, Spain, in 2006, 2007, and 2011, respectively.

He is currently a Post-Doctoral Researcher with the Department of Wireless Networks and Multimedia Services, Network and Service Architecture Group, Institut Telecom Sud Paris, Paris, France. He has published more than 20 papers in high-quality international conferences such as ACM CoNEXT, ACM MSWiM, IEEE ICC, IEEE ISCC, etc., and journals such as *Computer Networks*, IEEE COMMUNICATIONS LETTERS, *IEEE Communications Magazine*, *Sensors*, etc. His research is focused on the areas of social networking, peer-to-peer systems, and sensor networks.

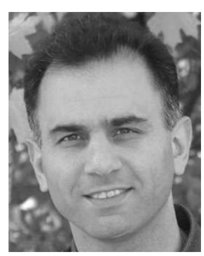
Dr. Cuevas was the recipient of the Best Paper Award in the ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (ACM MSWiM) 2010.

Sebastian Kaune received the Dipl. Inf. (M.S.) degree in computer science and Ph.D. degree in electrical engineering from the Technische Universität Darmstadt, Darmstadt, Germany, in 2006 and 2011, respectively.

In 2007, he joined the Multimedia Communications Lab (KOM), Technische Universität Darmstadt. Since 2012, he works as a consultant for IT process optimization in the chemical industry.

Carmen Guerrero received the Telecommunication Engineering degree from the Technical University of Madrid (UPM), Madrid, Spain, in 1994, and the Ph.D. degree in computer science from the Universidade da Coruña (UDC), A Coruña, Spain, in 1998.

Since 2003, she has been an Associate Professor with the Telematics Engineering Department, Universidad Carlos III de Madrid, Leganes, Spain. She has been involved in several national and international research projects related to green networking, future Internet, content distribution, and Internet measurements.



Reza Rejaie (M’00–SM’06) received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1991, and the M.S. and Ph.D. degrees in computer science from the University of Southern California, Los Angeles, in 1996 and 1999, respectively.

He is currently an Associate Professor with the University of Oregon, Eugene. From 1999 to 2002, he was a Senior Technical Staff Member with AT&T Labs—Research, Menlo Park, CA. His research interests include P2P networking, network measurement, online social networking, and multimedia networking.

Dr. Rejaie has been a Senior Member of the Association for Computing Machinery (ACM) since 2006. He received an NSF CAREER Award for his work on P2P streaming in 2005 and a European Union Marie Curie Fellowship in 2009–2010.