Content Publishing in BitTorrent

Shin So Eui
2012/08/22
Is Content Publishing in BitTorrent Altruistic or Profit-Driven?

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

Michal Kryczka  
Institute IMDEA Networks and  
Univ. Carlos III de Madrid  
michal.kryczka@imdea.org

Angel Cuevas  
Univ. Carlos III de Madrid  
acrumin@it.uc3m.es

Sebastian Kaune  
TU Darmstadt  
kaune@kom.tu-darmstadt.de

Carmen Guerrero  
Univ. Carlos III de Madrid  
guerrero@it.uc3m.es

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

Co-NEXT(Conference on emerging Networking EXperiments and Technologies)  
'10 Proceedings of the 6th International Conference
Contents

1. Introduction
2. Paper’s Point
3. Measurement Methodology
4. Identifying Major Publishers
5. Signature Of Major Publishers
6. Incentive of Major Publishers
7. Other Beneficiaries in The BitTorrent Marketplace
8. Conclusion
1. Introduction

BitTorrent is the most popular P2P content delivery application

The availability of valuable content without any cost for the consumers

- Most of major publishers dedicate their resources for publishing content and consume few or no published content by others.

- Their level of content publication and consumption is very imbalanced.

An Important Question

The incentive of publisher who make these content available through BitTorrent portals

HOW!!(incentive)
WHY!!(publishing contents)
2. Paper’s Point

This Paper’s Point is an Analysis of SOCIO-ECONOMIC ASPECT

1) Identifying the content publishers

Two different profiles

Fake publisher
Publishing a large amount of fake files (antipiracy agencies and malicious publishers)

Top publisher
Publishing a large number of proper (often copyrighted) content.

2) Investigating the main incentives of major (non-fake) publishers and classifying them

1. Private BitTorrent Portals that offer certain services and receive financial gain through ads, donations and fees

2. Promoting web that leverage published content at BitTorrent portals to attract users to their own web site for financial gain

3. Altruistic Major Publishers
3. Measurement Methodology

- This paper first briefly describes the required **background** on how a user joins a BitTorrent swarm.

**Background**

**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.**

(i) the number of seeders and leechers that are currently connected to the swarm

(ii) N (typically 50) random IP addresses of participating peers in the swarm

**Identifying Initial Publisher**

Authors often contact the tracker shortly after the birth of the associated swam when the number of participating peers is likely to be small and the initial publisher (i.e. seeder) is one of them.

**To Retrieve**

1) **The IP address of all participating peers** and 2) **The current number of seeders in the swarm**

- we obtain the bitfield of available pieces at individual peers to identify the seeder.
- Otherwise, reliably identifying the initial seeder is difficult because there are more than one seeder or the number of participating peers is large.
4. Identifying Major Publishers (1/2)

Dataset

<table>
<thead>
<tr>
<th>Portal</th>
<th>Start</th>
<th>End</th>
<th>#Torrents</th>
<th>#IP addresses</th>
</tr>
</thead>
<tbody>
<tr>
<td>mn08</td>
<td>Mininova</td>
<td>09-Dec-08</td>
<td>16-Jan-09</td>
<td>/20.3K</td>
</tr>
<tr>
<td>pb09</td>
<td>Pirate Bay</td>
<td>28-Nov-09</td>
<td>18-Dec-09</td>
<td>23.2K/10.4K</td>
</tr>
<tr>
<td>pb10</td>
<td>Pirate Bay</td>
<td>06-Apr-10</td>
<td>05-May-10</td>
<td>38.4K/14.6K</td>
</tr>
</tbody>
</table>

Table 1 shows the main features of our three datasets (from Mininova and from the Pirate Bay)

- The number of torrents for which we identified the initial publisher (username/IP address).
- The total number of discovered IP addresses associated for all the monitored swarms.
- Dataset mn08 does not contain the username of initial publishers and authors focused on pb10 for paper’s detailed analysis.

4.1 Skewness of Contribution

- Top 3% of the BitTorrent publishers contribute roughly 40% of the published content.
- 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 5 files (80%).

4.2 Publisher’ ISPs

1) A Commercial ISP
   ex) Comcast

2) A Hosting Provider
   ex) OVH

Figure 1: Percentage of content published by the top x% publishers.
4. Identifying Major Publishers (2/2)

4.2 Publisher’ ISPs

<table>
<thead>
<tr>
<th></th>
<th>Published torrents</th>
<th># IP addr</th>
<th># /16 IP</th>
<th># Geo Loc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>OVH (mn08)</td>
<td>2766</td>
<td>164</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Comcast (mn08)</td>
<td>976</td>
<td>675</td>
<td>269</td>
<td>400</td>
</tr>
<tr>
<td>OVH (pb09)</td>
<td>2577</td>
<td>78</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Comcast (pb09)</td>
<td>382</td>
<td>198</td>
<td>143</td>
<td>129</td>
</tr>
<tr>
<td>OVH (pb10)</td>
<td>2213</td>
<td>92</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Comcast (pb10)</td>
<td>408</td>
<td>185</td>
<td>139</td>
<td>147</td>
</tr>
</tbody>
</table>

Table 3: Characteristics of all OVH and Comcast publishers in mn08, pb09 and pb10.

The aggregate contribution of each publisher at OVH is on average a few times larger than Comcast publishers.

The published content by Comcast publishers comes from a large number of typical altruistic users where each one publishes a small number of files 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.

4.3 A Closer Look at Major Publisher

The top-100 IP addresses

- 45% a unique username
- 55% a large number of usernames

The top-100 usernames

- 34% Hosting Provider
- 24% A Single Commercial ISP
- 17% Different Commercial ISPs
- 75% multiple IPs
- 25% a single IP
- 25% Without fake publisher’s content
- 37% upload
- 50% download
5. Signature Of Major Publishers (1/2)

Incentives characteristics of major publishers in dataset

all publishers (labeled as “All”),
all fake publishers (labeled as “Fake”)
all top-100(non-fake) publishers regardless of their ISPs (labeled as “Top”).
ISPs into hosting providers (labeled as “Top-HP”)
ISPs into commercial ISPs (labeled as “Top-CI”)

5.1 Content Type

• Video files constitute a significant fraction of published files across most groups with some important differences.

• Fake publishers primarily focus on Videos and Software content.

5.2 Content Popularity

• Fake publishers' content is the most unpopular among the target groups.
• Top publishers are responsible for a larger fraction of popular torrents.
5. Signature Of Major Publishers (2/2)

5.3 Seeding Behavior

**Average Seeding Time:** The seeding time for fake publishers is significantly longer than publishers in other groups.

**Average number of Parallel Torrents:** fake publishers typically publish a large number of torrents and other users do not help them for seeding.

**Aggregated Session Time:** As expected fake publishers present the longest aggregated session time due to their obligation to continuously seed their content to keep them alive.
6. Inventive of Major Publishers (1/2)

- **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

6.1 Classifying Publishers

**Private BitTorrent Trackers (top of 25%)**

Three different ways to gain financial profit:
(i) **posting advertisement** in their web sites, (ii) **seeking donations from visitors** to continue their basic service, and (iii) **collecting a fee for VIP access** that allows the client to download any content without requiring any kind of seeding ratio.

**Promoting Web Sites (top of 23%)**

Most of these publishers (70%), specifically those that are running a hosting image web site, publish only porn content.
Inspection of the associated hosting image web sites revealed that they store adult pictures.

**Altruistic Publisher (top of 52%)**

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.
6. Inventive of Major Publishers (2/2)

6.2 Longitudinal View of Major Publishers

(i) Publisher Lifetime which represents the number of days between the first and the last appearance of the publisher in the Pirate Bay portal.

(ii) Average Publishing Rate that indicates the average number of published content per day during their lifetime.

<table>
<thead>
<tr>
<th>Class</th>
<th>Lifetime</th>
<th>Avg. Publishing Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private Portals</td>
<td>63/466/1816</td>
<td>0.57/11.43/79.91</td>
</tr>
<tr>
<td>Promoting Web sites</td>
<td>50/459/1899</td>
<td>0.38/4.31/18.98</td>
</tr>
<tr>
<td>Altruistic</td>
<td>10/376/1899</td>
<td>0.10/3.80/23.67</td>
</tr>
</tbody>
</table>

Table 4: Lifetime and Avg. 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

6.3 Estimating Publishers’ Income

The goal of half of the top publishers is to attract users to their own web sites. Most of these publishers seem to generate income at least by posting ads in their web sites.

(i) average value of the web site
(ii) average daily income of the web site
(iii) average daily visits to the web site
7. Other Beneficiaries in The BitTorrent Marketplace

Major Public BitTorrent Portals

- The main advantage of these major portals is that they offer a reliable service (e.g. they rapidly react to remove fake or infected content).
- These major portals are one of the key players of the BitTorrent Ecosystem that brings substantial financial profit.

Hosting Providers

Heavy seeding activity performed by some publishers requires significant resources

Ad Companies

- They charge their customers for this service and part of this income is forwarded to the web sites where the ads have been posted.
- Therefore, ad companies look for popular web sites where to put ads for their costumers.
8. Conclusion

In this paper we studied the content publishing activity in BitTorrent from a socio-economic perspective.

The incentives of major publishers and identified the following key characteristics:

first, antipiracy agencies and malicious users perform a systematic poisoning index attack over major BitTorrent portals in order to obstruct download of copyrighted content and to spread malware, respectively.

Second, 37% of the (non-fake) published content is published by a small fraction of users that serve 54% of the (non-fake files) downloads. Our evidence suggests that these publishers have financial incentives for posting these contents on BitTorrent portals.

The removal of these financial-driven publishers (e.g. by antipiracy actions) may significantly affect the popularity of these portals as well as the whole BitTorrent ecosystem.
Thank you for listening & Q&A