Is Content Publishing in BitTorrent Altruistic or Profit-Driven?

Rubén Cuevas (UC3M), Michal Kryczka (Institute IMDEA Networks), Angel Cuevas (UC3M), Sebastian Kaune (TU Darmstadt), Carmen Guerrero (UC3M) and Reza Rejaie (Univ. Oregon)

Rubén Cuevas
rcuevas@it.uc3m.es
University Carlos III of Madrid
Outline

1. Motivation
2. Data Collection
3. Identifying Major Publishers
4. Incentives of Major Publishers
5. Conclusion
Outline

1. Motivation

2. Data Collection

3. Identifying Major Publishers

4. Incentives of Major Publishers

5. Conclusion
BT is massively used and responsible of a large portion of the Internet Traffic. Its success is due to the availability of (often copyrighted) valuable/popular content at very low cost.

Publishing content implies:
- Spend Resources (cpu, bw, …)
- Legal implications (fines, even prison)

What are the main incentives of content publishers in BitTorrent?

1. Altruism (good citizens)?
2. Profit-driven incentive?
3. Any other incentive?
Outline

1. Motivation
2. Data Collection
3. Identifying Major Publishers
4. Incentives of Major Publishers
5. Conclusion
Data Collection

• Content Publisher can be identified by
  - **USERNAME** used to register the torrent in a BitTorrent Portal
  - **IP address** used for initially seed the content in the swarm

• We crawl BitTorrent Portals
  - To retrieve the Content Publisher’s username (easy)

• We crawl BitTorrent Trackers:
  - To retrieve the Content Publisher’s IP address (difficult)
  - To retrieve the IP addresses of the downloaders of a given content
## Our Dataset

<table>
<thead>
<tr>
<th>Dataset name</th>
<th>BitTorrent Portal</th>
<th>Start Date</th>
<th>End Date</th>
<th># Torrrents (username/ IP addr)</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>- / 20.8K</td>
<td>8.2M</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>
<td>52.9K</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>
<td>27.3M</td>
</tr>
</tbody>
</table>
1. Motivation

2. Data Collection

3. Identifying Major Publishers

4. Incentives of Major Publishers

5. Conclusion
40% of the content is published by 3% (~100) of the publishers.
Identifying major publishers

- Top-100 publishers based on IP addresses
  - 55% use a unique username
  - 45% use a large number of usernames -> “Fake”
  - Hacked or random-name usernames
  - Fake content -> PB removes the username account
  - 25% usernames, 30% content, 25% downloads

- Top-100 publishers based on usernames
  - 25% use a unique IP
  - If we remove the fake usernames -> “Top”
  - 37% content and 50% downloads

“Top” + “Fake” publishers are responsible of
66% of the content
75% of the downloads
Signature of Content Publishers

• Do major publishers have any distinguishing features (i.e. signature)?

• We examine the following characteristics
  - Type of published content
  - Popularity of published content
  - Availability and Seeding behavior
    • Aggregated Session Time per Publisher
"Fake" publishers’ content is the most unpopular:
- Portals remove identified fake content
- Users quickly report fake content in forums

The median popularity of “Top” publishers’ torrents is 7 times higher than a typical user (“All”).
Seeding behavior of content publishers
Aggregated Session Time per Publisher

Nobody helps on seeding fake content.

“Top” publishers much more available than regular publishers (more contents seed during longer periods)
Outline

1. Motivation

2. Data Collection

3. Identifying Major Publishers

4. Incentives of Major Publishers

5. Conclusion
Incentives of “Fake” publishers

- They publish files with catchy titles – e.g. recently released Hollywood movies
- We download few files from them and found two profiles.
- Anti-piracy agencies publish fake version of the content to be protected – loops with few scenes or anti-piracy ads
- Malicious users try to spread malware – The content point to an specific player to be reproduced that is a malware

Is Content Publishing in BitTorrent Altruistic or Profit Driven?
ACM CoNEXT 2010, Philadelphia
Incentives of “Top” publishers

Data Collection

• We emulate the experience of a user by downloading a few files published by each “Top” publisher looking for a promoted URL

  – Name of the downloaded file
  – Torrent webpage’s textbox
  – Name of a text file displayed when opening the torrent

• Publisher’s username

  – Username = promoted URL (e.g. ultrorrents)

• Business Service

  – Type of service offered at the promoted URL

---

Is Content Publishing in BitTorrent Altruistic or Profit Driven?

ACM CoNEXT 2010, Philadelphia
Incentives of “Top” publishers Profiles

- **Private BT Portals/Trackers (profit-driven)**
  - 18% of the content, 29% of the downloads
  - Financial profit: ads, donations and VIP access
  - Video, audio and software

- **promoted Web Sites (profit-driven)**
  - 8% of the content, 11% of the downloads
  - Mostly hosting (adult) images portals that publish porn content in the Pirate Bay
  - Financial profit: ads

- **Altruistic (major) Publishers**
  - 11.5% of the content, 11.5% of the downloads
  - Light content (music and e-books)
  - Detailed description, asking for help on seeding
Estimating Publishers’ Income of profit-driven publishers

- Promoted Web Sites are fairly profitable and visible

Few of the profit-driven publishers are associated to very valuable web sites (valued up to $ millions)

<table>
<thead>
<tr>
<th></th>
<th>Web Site Value ($)</th>
<th>Web Site Daily Income ($)</th>
<th>Web Site Daily Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min / Median / Max</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private BT Portals</td>
<td>1K/33K/2.8M</td>
<td>1/55/3.7K</td>
<td>74/21K/1.4M</td>
</tr>
<tr>
<td>Promoted Web Sites</td>
<td>24/22K/1.8M</td>
<td>1/51/1.9K</td>
<td>7/22K/772K</td>
</tr>
</tbody>
</table>
Outline

1. Motivation
2. Data Collection
3. Identifying Major Publishers
4. Incentives of Major Publishers
5. Conclusion
Conclusion

- 67% of the published content and 75% of the downloads depend on few publishers.

The removal of the financial-driven publishers (e.g. by antipiracy legal actions) may severely affect the popularity of BitTorrent. If this happens:

http://bittorrentcontentpublishers.netcom.it.uc3m.es

“Will BitTorrent survive as the most popular file-sharing application without these financial-driven publishers?”

87% of non-fake content and 84% non-fake downloads
Is Content Publishing in BitTorrent Altruistic or Profit-Driven?

Rubén Cuevas (UC3M), Michal Kryczka (IMDEA Networks), Angel Cuevas (UC3M), Sebastian Kaune (TU Darmstadt), Carmen Guerrero (UC3M) and Reza Rejaie (Univ. Oregon)

Thanks for your attention!

Rubén Cuevas
rcuevas@it.uc3m.es
University Carlos III of Madrid

Is Content Publishing in BitTorrent Altruistic or Profit Driven?
ACM CoNEXT 2010, Philadelphia