BitTorrent ecosystem

BitTorrent is the most successful peer-to-peer file-sharing application. Indeed, it is responsible of a major portion of the Internet traffic share and is daily used by hundred of millions of users.
We present the techniques which allow to evaluate the performance and the demographics aspects of BitTorrent.
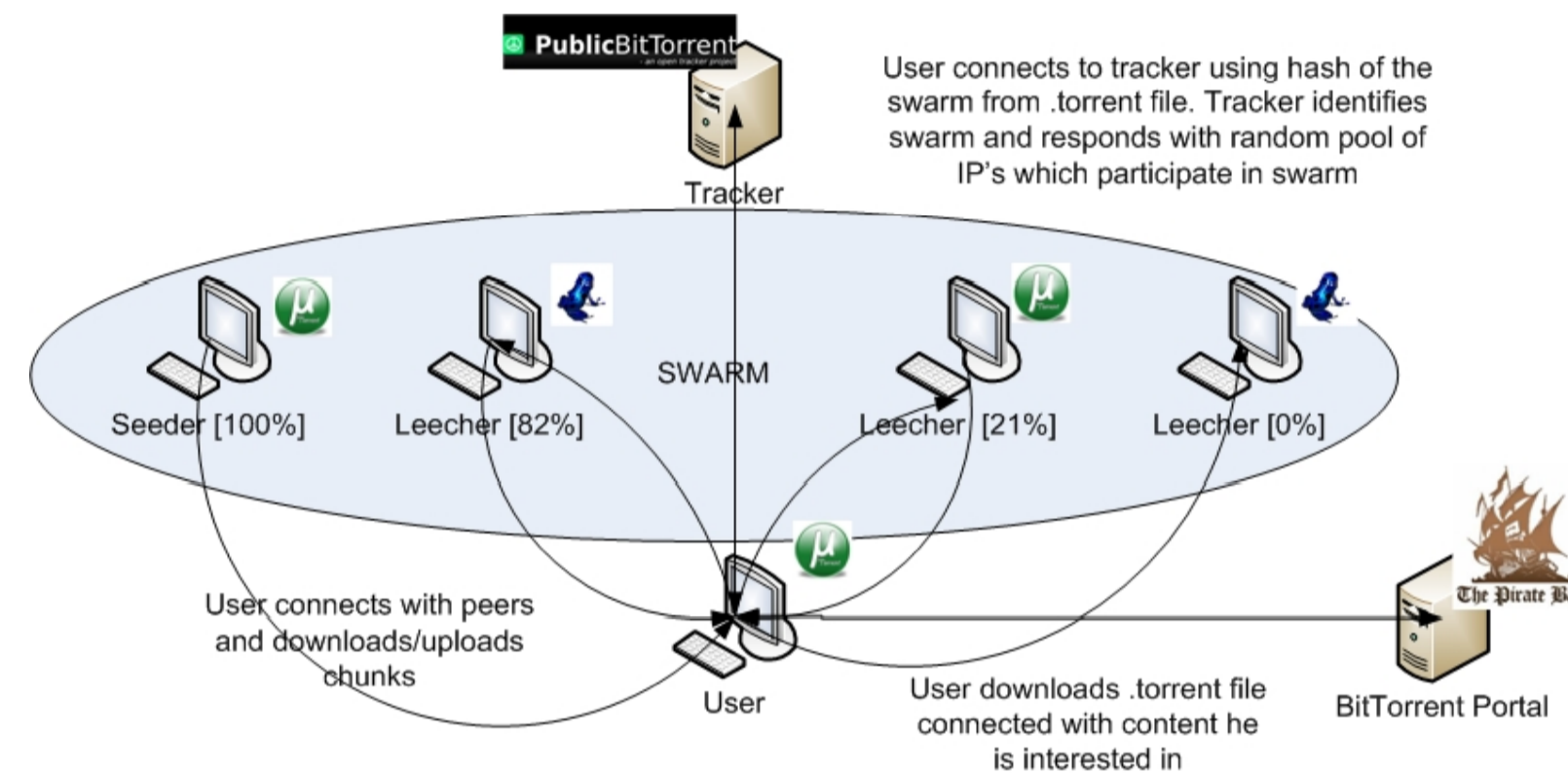
BitTorent Ecosystem

Webpage Crawling

Once know the id assigned to a given torrent in The Pirate Bay, we just need to access its web-page and parse it (using an html-parser) to retrieve the torrent information. We can set the range manually or use RSS mechanism to crawl forwardly.
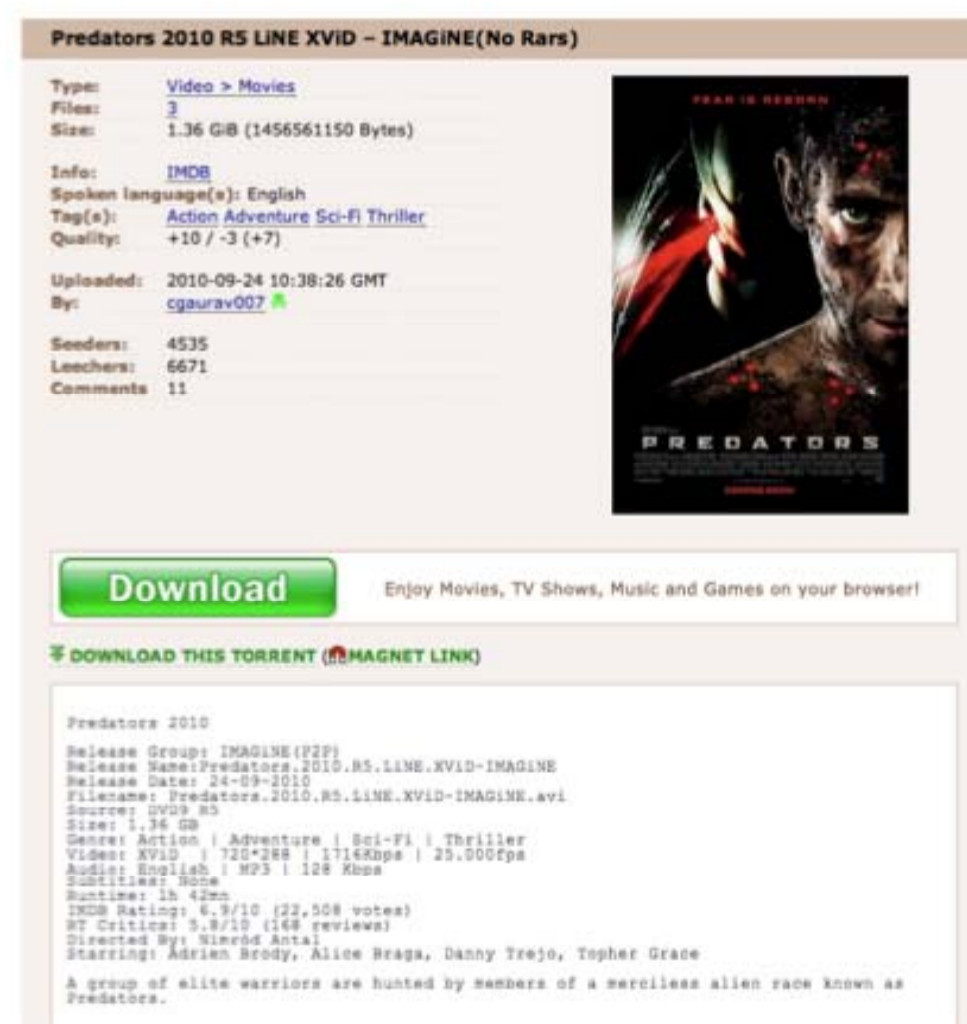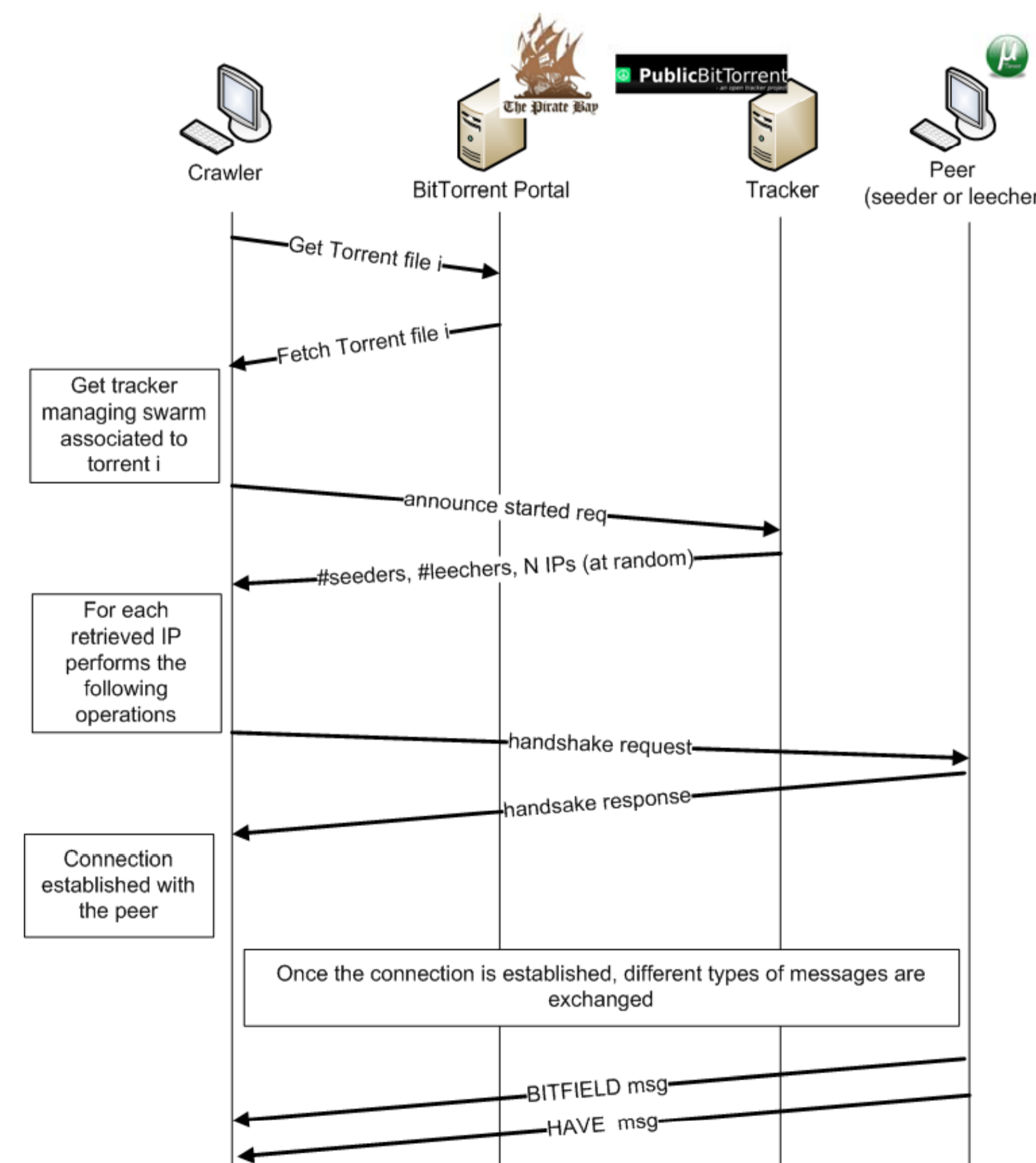

Possible interesting results:
• Content Popularity Distribution
• Distribution of number of published content per category and subcategory
• Torrents Publishing Rate per date

Crawling (tracker + peers)

By systematical querying the tracker of particular torrent you can obtain IP addresses of peers participating in swarm. Moreover connecting to these peers makes possible to obtain such information like client the peer uses, downloaded chunks, his routing table etc.

Possible interesting results:
• The Distribution of clients per country or ISP
• Heavy Hitters
• Peers' Arrival Rate and Session Time
• Peer's performance
• Chunk distribution

Challenges

Several challenges to overcome exist:
• Peer Identification
  Peers does not have permanent identification. IP can be dynamically assigned, so it is hard to follow a peer across multiple session

• Crawler's IP address banned by the Tracker
  Tracker can monitor the traffic and once suspicious behaviour is observed (e.g. querying with high rate) IP of crawler (or even whole /24 mask) can be banned

• Crawler's IP address blacklisted by the client
  When connecting to Client, Crawler does not offer any chunk to a peer. This can lead to blacklist the IP of the Crawler for part of the time

• Completeness of data
  It is impossible to connect to peers behind NAT, we need to wait till they will try to connect with us

References

R.Cuevas,M.Kryczka,A.Cuevas,C.Guerrero,A.Azcorra
Measuring BitTorrent Ecosystem: Techniques, Tips and Tricks.
Work-in-progress

R.Cuevas,M.Kryczka,A.Cuevas,S.Kaune,C.Guerrero, R.Rejaie
Is Content Publishing in BitTorrent Altruistic or Profit-Driven?
In ACM CoNEXT 2010, 2010.