

Xunlei: Peer-Assisted Download Acceleration on a Massive Scale

Prithula Dhungel¹, Keith W. Ross¹, Moritz Steiner², Ye Tian³, Xiaojun Hei⁴

¹ Polytechnic Institute of NYU, Brooklyn, NY, USA

² Bell Labs, Alcatel-Lucent, Holmdel, NJ, USA

³ University of Science and Technology of China, Hefei, China

⁴ Huazhong University of Science and Technology, Hubei, China

Abstract. We take a close look at Xunlei, an enormously popular download acceleration application in China. Xunlei forms a complex ecosystem, with Xunlei peers extensively interacting with independent HTTP and FTP servers, cyberlockers (such as megaupload and hotfile), the BitTorrent and eDonkey file-sharing systems, as well as with other Xunlei peers. After performing a protocol analysis on Xunlei, we develop a comprehensive measurement infrastructure, enabling us to gain new insights into the scale of content, swarm sizes, and several unique characteristics of the system mechanisms in Xunlei.

1 Introduction

In this paper, we take a close look at Xunlei, an enormously popular download acceleration application in China. Xunlei is largely confined to China and has received relatively little attention in the research community to date. It is an interesting application for many reasons. First, it employs several interesting innovations, many of which we will explore in this measurement study. Second, it forms a complex ecosystem, extensively interacting with independent servers, cyberlockers (such as megaupload and hotfile), and two major peer-to-peer (P2P) file sharing systems. Third, as demonstrated in our study, it is deployed on a massive scale. To the best of our knowledge, only a few preliminary studies of Xunlei have been carried out to date, focusing on the protocols used for transferring data among peers [4, 5].

When installing the *Xunlei client*, a plug-in is also installed into Internet Explorer. With this plug-in installed, when a user downloads a file from within the browser, the Xunlei client takes over the download. Xunlei can be used to download various types of files including video files, pdf files, executables, and even email attachments. The Xunlei client supports multiple protocols for file download, including HTTP, FTP, RTSP, and MMS. The Xunlei client also serves as a BitTorrent (BT) and eDonkey client. It was reported recently that Xunlei has recently become the most popular BitTorrent client [2].

Xunlei further provides peer-assisted and server-assisted download acceleration. Whenever the Xunlei client is instructed to download a file from an HTTP server, from an FTP server, from the BT system, or from the eDonkey system,

the Xunlei client contacts Xunlei’s centralized tracker, which provides the client a list of other Xunlei peers that share this file. The client can then download portions of the file from other Xunlei peers while also downloading portions from the original source. Xunlei clients communicate with the tracker and communicate with each other using proprietary Xunlei protocols. Moreover, in addition to providing a list of Xunlei peers that have the file, the tracker also provides servers (HTTP, FTP) that have the file. To the best of our knowledge, Xunlei is the only widely-deployed download manager that provides both *peer and server* assistance when downloading files.

To assist users in locating files, Xunlei provides a search engine, *gougou.com*, which indexes a subset of the content indexed by the Xunlei tracker system. For each Gougou indexed file, there is a pointer to a download link (e.g., to an HTTP server, an FTP server, or a BT infohash) for initiating download with the Xunlei client. Recently, the Xunlei network also started a VoD service, known as *Xunlei Kankan*, which serves thousands of movies and TV shows. To view Kankan videos, a user needs to install the proprietary *Kankan client*. As with Xunlei file downloads, Kankan video distribution is peer-assisted. An extended version of this paper is available as a technical report [3]. Our contributions in this paper include:

1. Xunlei uses encrypted proprietary protocols for communication between Xunlei clients, between Xunlei client and Xunlei tracker, and between Xunlei client and Xunlei’s proprietary distributed hash table (DHT). We performed a protocol analysis, uncovering Xunlei protocol message formats, and formats for Xunlei peer IDs.
2. Based on our understanding of the Xunlei protocols, we developed a measurement infrastructure for the Xunlei ecosystem. Our measurement platform consists of a Xunlei DHT crawler, a Xunlei tracker crawler, and several website crawlers.
3. Given that the Xunlei tracker tracks content emanating from many different sources and protocols, we investigated the scale of content being tracked. We compared the amount of content being tracked by Xunlei with the amount of content being tracked by the (currently) most popular BT tracker. We found swarm sizes in Xunlei and BT to be of comparable size.
4. We performed an analysis of whether content flows from BT to HTTP and FTP servers via Xunlei. We found that such flows are common for popular BT content, and often the files are deposited into cyberlocker sites. We observed that Xunlei is also providing peer-assisted download acceleration for large email attachments hosted in mail servers.
5. By crawling the Xunlei trackers, we determined the swarm sizes for the VoD movies available from Kankan. We found Kankan swarm sizes to be much larger than the largest swarms in BitTorrent.

2 New Insights into Xunlei

Xunlei uses a proprietary, closed source, protocol. In order to develop measurement tools for the Xunlei ecosystem, we carried out a laborious analysis of the

Xunlei protocols, including determining the syntax and semantics of many of the protocol messages, as well as the formats for Xunlei peer IDs. This was a particularly challenging task since most of the fields in the Xunlei packets are encrypted. In this process, we discovered that Xunlei clients also join two separate DHTs: its own proprietary DHT and also the Mainline BitTorrent DHT. As part of our packet analysis, we were also able to understand how the Xunlei DHT operates and many of the DHT message formats. To this end, we have also understood the details of the encryption/decryption algorithms used. Xunlei uses AES in ECB mode for encrypting messages exchanged between its entities. The 64-bit key for each message is pre-pended to the message itself. To understand the details of the Xunlei protocol, we used Wireshark to passively monitor the traffic flowing in and out of a Xunlei client in our university network, and then later studied this data.

The Xunlei tracker plays a central role in the Xunlei ecosystem. It not only tracks a huge number of files, but also tracks files originating from many different sources, including cyberlockers, FTP servers, mail servers. We have found that whenever a Xunlei peer downloads a file (or a portion thereof) from some source, it notifies the tracker. In this manner, when another Xunlei client wants to download the file, the tracker can provide all known sources, including peers and servers.

We have observed that each peer in Xunlei uses different identifiers for itself when joining different networks. For example, it uses a 16-byte identifier when joining the KAD network, a 20-byte identifier in the Xunlei DHT, a 20-byte identifier for BitTorrent, and a 16-byte unique identifier when registering itself with the Xunlei central trackers when downloading files. We refer to this last type of identifier as the *Xunlei ID* for the peer. Its first 12 bytes correspond to the hexadecimal equivalent of the MAC address of the machine the client is running on. Therefore, each Xunlei client can be uniquely identified by its Xunlei ID over extended periods of time.

Different types of sources use different types of identifiers for the same file: Servers use URLs, BT uses a 20-byte infohash of the file, and eDonkey uses the 16-byte eDonkey hash extracted from the ed2k link. In order for the tracker to provide cross-protocol sources for a particular request, the Xunlei client constructs an internal hash for each file it has been downloaded, and then sends this hash to a tracker, along with the identifiers of the sources from which it downloaded the file. The tracker most likely has a hash table, with the internal hash being the key, and a list of all sources that are known to have the file, with the identifier type being source specific. These sources can include HTTP, FTP, RTSP, and MMS URLs, a BT infohash, an eDonkey hash, and the Xunlei IDs of the peers holding the file.

One can view Xunlei as an effort to consolidate the many different ways of exchanging content on the Internet. The user does not need to be concerned about whether the content is hosted by cyberlockers, available on some P2P file-sharing network, or on traditional web servers. The underlying infrastructure is abstracted and made compatible on a chunk level.

2.1 Multi-protocol Support

As an example of Xunlei's cross-protocol operation, consider a file that is present in an HTTP server and also within BT. Suppose initially the tracker is not aware of this file. When a Xunlei client downloads the file from BT, it calculates the internal hash of the file and informs the tracker of the mapping between the BT infohash and the internal hash. When another Xunlei client downloads the same file directly from the server, it calculates the (same) internal hash for the file and informs the tracker of the mapping between the HTTP URL and the internal hash.

In this manner, the Xunlei tracker knows that the BT infohash and the HTTP URL correspond to the same file. Thereafter, if a Xunlei client initiates a download with the BT infohash, the Xunlei tracker can provide it with the HTTP server link and the Xunlei peers that have copies of the file. We performed several simple tests for verification. *For example, when using Xunlei to download a particular BT file, only 4% of the file came from BT, the remainder of the file came from an HTTP server (74%) and from Xunlei peers using the Xunlei protocol (22%).*

2.2 Bandwidth Leeching

Since the Xunlei tracker provides the Xunlei client with a list of servers currently hosting a file, when a user initiates download for a file from gougou.com (or any other location), the Xunlei client can often download much of the content in the file from a non-Xunlei server, which typically use advertising to generate revenue. Xunlei thus negatively impacts their profits, since users are downloading from them without eyeballing their websites. This phenomenon is referred to as *bandwidth leeching*. Xunlei allows, however, such servers to receive compensation by joining the *Xunlei Union*. We also note that Xunlei provides bandwidth assistance to non-Xunlei servers and sources. For example, when a Xunlei client initiates a download directly from an HTTP server (for example, a cyberlocker such as megaupload), if other Xunlei peers have already downloaded the file (or are currently downloading it), then those peers can assist the server. As we will discuss in the body of the paper, even the downloading of e-mail attaches can benefit from Xunlei's bandwidth assistance.

As mentioned earlier, the Xunlei client acts as a BT client during BT downloads. However, unlike other BT clients, it does not allow a user to create a .torrent file. The user instead would have to use a traditional BT client to do that. However, once the .torrent file (and then the file corresponding to the .torrent) is downloaded using a Xunlei client, the file gets tracked by the Xunlei tracker and hence enters the Xunlei domain. After this, when a Xunlei client begins to download the BT file, it receives peer-assistance from Xunlei peers (which may no longer be active in the swarm but nevertheless have a copy of the file).

2.3 Tracker and Client Protocol

By monitoring and analyzing the traffic at a local Xunlei client, we were able to identify the details of messages exchanged by the client with the Xunlei trackers when requesting the resource list for any given file. We now discuss some of the more important observations we made about the protocol.

When downloading a file using a resource link (e.g., made available by gougou.com), the Xunlei client sends a message with the link to the central tracker, which in turn returns two 20-byte hash values and a single 8-byte code corresponding to the file. These hash values and the code are then used to request the peer and server resource lists for the file. The two hash values and the code for a file can also be obtained from the trackers by sending the unique 20-byte identifier for the file. For files indexed in gougou.com, the identifier is also available in its gougou page.

For BitTorrent files, the Xunlei client uses the infohash of the file as the 20-byte identifier in the message sent to the tracker. For eDonkey files, the 20 byte identifier is obtained from the 16 byte hash extracted from the ed2k link along with the file size. When the Xunlei client sends to the tracker an identifier that is *not* currently being tracked by Xunlei, the tracker does not return the two hash values and the code. Through out the paper, we consider a file as being “tracked” by Xunlei if the trackers return the two hash values and the code for the file. This is the technique that we use in Section 3.1 to evaluate the fraction of BitTorrent and eDonkey files tracked by Xunlei.

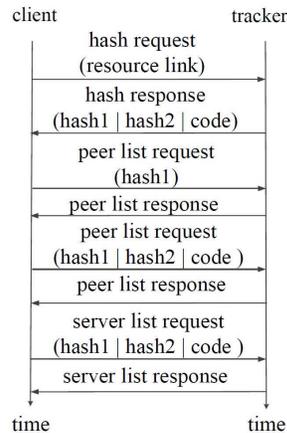


Fig. 1. Message exchange between Xunlei client and trackers when requesting resource list.

Country	Peers	Population	Ratio
Mainland China	5,261,157	1,330.1	3,955
Taiwan	57,852	23.0	2,512
Hong Kong	20,222	7.1	2,852
Malaysia	7,340	26.2	280
Korea	7,043	48.7	144
Singapore	4,865	4.7	1,034
Japan	4,326	126.8	34
United States	4,237	310.2	13
Macau	2,995	0.6	5,273
Italy	1,802	58.1	31
United Kingdom	1,497	62.3	24
Canada	1,150	33.8	34
Australia	813	21.2	38
Venezuela	616	27.2	22
France	470	64.8	7

Fig. 2. Country Level Distribution of Peers

3 Measuring the Scale of Xunlei

We used the results from protocol analysis to develop a measurement infrastructure for the Xunlei network. As shown in Figure 3, our measurement platform consists of three major components: a Xunlei DHT crawler, a Xunlei tracker crawler, and several website crawlers.

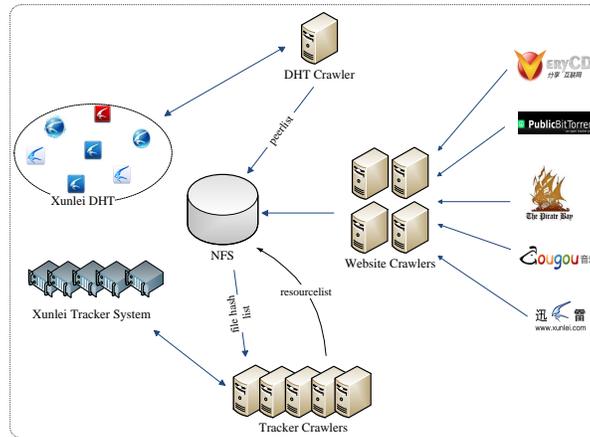


Fig. 3. Measurement Platform

Our *Xunlei DHT Crawler* recursively crawls the routing tables of currently online Xunlei peers to enumerate the list of Xunlei peers in the DHT. For a given 20-byte hash, the crawler works by contacting a Xunlei peer, and asking that peer for the list of peers in its routing table that are close to the hash. After receiving new peers, the crawler again contacts these peers to get more new peers and so on. Given an identifier for a particular file (URL, BT or eDonkey identifier), the *Xunlei Tracker Crawler* continuously contacts the Xunlei central tracker infrastructure to get the list of Xunlei peers that have the file. The tracker crawler also obtains a list of HTTP and FTP servers hosting the file.

We developed 4 different website crawlers and web page parsers customized for each website. The *Gougou Crawler* crawls the movie database in the gougou website. For each indexed movie, it extracts the download link from its movie information page and passes the links to the Tracker Crawler for gathering peer lists and server lists. The *PirateBay Crawler* monitors the Pirate Bay website for the latest TV shows uploaded to the website; immediately after a TV show is uploaded, it extracts the corresponding BT infohash for the TV show, and passes it to the Tracker Crawler to gather the Xunlei peers and HTTP/FTP servers.

The *Public BitTorrent Crawler* downloads the list of BT infohashes being tracked by the public BT tracker site publicbt.com - the largest BT tracker during the time of our experiments. Finally, the *Verycd Crawler* extracts the ed2k links for the eDonkey files indexed by the popular Chinese eDonkey website verycd.com.

3.1 Scale of Content Indexed

Given that the Xunlei tracker is tracking content originating from many different sources and protocols, it is of interest to estimate the scale of the tracker, that is, the number of files being tracked by Xunlei. In particular, how does the scale

of the China-centric Xunlei tracker compare with the 2.9 million files (see below) tracked by the largest BT tracker, which targets a culturally diverse user base?

One approach to getting a handle on the Xunlei content scale is to determine the fraction of the content in BT and in eDonkey that is being tracked by Xunlei. To this end, for BT we used the Public BitTorrent Crawler to get the list of 2,920,045 BT infohashes being tracked by the public BT tracker. For eDonkey, we used the Verycd Crawler to extract the hashes being indexed in the verycd.com website. We extracted 384,612 different file pages indexed by verycd.com, corresponding to a total of 1,064,525 different ed2k hashes (each file can have multiple parts, each of which has a distinct hash value). These BT and eDonkey hashes were then passed to the Tracker Crawler that determines which of these are being tracked by Xunlei trackers. *We found that 90.3% of the files indexed by verycd.com and 47.4% of the files indexed by Public BitTorrent are being tracked by Xunlei. Thus, the scale of Xunlei's tracking is at least on the order of BT, perhaps significantly larger (see below).*

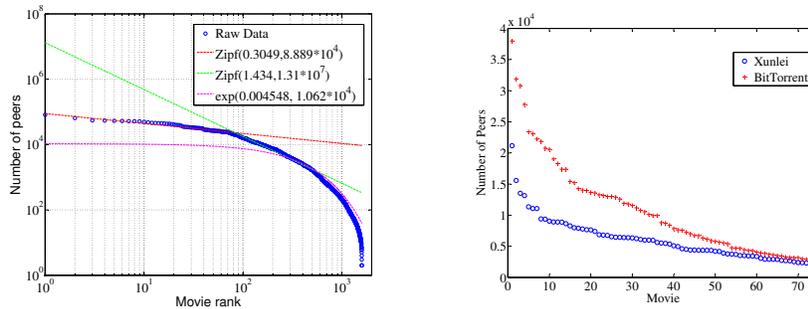
Another approach to getting a handle on the Xunlei content scale is to crawl gougou.com and determine the number of files indexed there. Using our Gougou Crawler, we found that gougou indexes 1,092,114 distinct files in the movie category. There is also over one million distinct files in the other Xunlei categories. So the Xunlei tracker is tracking over 2 million files. But does the Xunlei tracker track files that are not indexed by Gougou? Perhaps the files in Gougou only represent the tip of the iceberg? To this end, we observed that approximately 110,000 Gougou movie files have sources in BT. On the other hand, of the 2.9 million files indexed by Public BitTorrent tracker, roughly 600,000 are movie files. Of these 600,000 movie files indexed by Public BitTorrent tracker, roughly 280,000 are indexed by the Xunlei tracker (see the 47.4% overlap result above). Therefore, the number of BT movie files the Xunlei tracker is tracking at least 2.5 times more than indexed by the Gougou. Hence, the Xunlei tracker is indeed tracking significantly more content than indexed in Gougou.

3.2 Xunlei Swarm Scales

The gougou.com website provides a list of all versions of the top 100 movies in Xunlei. Using our Gougou Crawler, we extracted the download links for each of these movie versions and passed them to the *TrackerCrawler* component. Figure 4(b) shows the swarm size distribution of the top 75 movies in Xunlei. For a comparison, we also plot the total swarm sizes for top 75 movies in Pirate Bay. For calculating the total swarm size for each movie title in Pirate Bay, we used carefully (manually) chosen strings as inputs to the Pirate Bay movie search engine and obtained the list of versions for each of the top 75 movie titles and added up the swarm sizes for all the versions. *Although the largest movie swarm in Xunlei is smaller than that for Pirate Bay, we see that the swarm sizes start to converge to similar values after the 60th most popular movie.* We also crawled the trackers for lists of peers watching movies and TV show episodes served by Kankan. Figure 4(a) shows the snapshot of distribution of swarm sizes for all 1,591 movies available for VoD. (We performed a number of tracker crawls from

Dec 2010 to Jan 2011. In Figures 4(b), 4(a), we plot the results for crawls that are close representatives of other crawls.) *Kankan movie swarms are as large as 69,573, much larger than the largest movie swarms in BT.* By crawling Kankan TV shows for a week, we collected 7,358,040 unique users. Recall that Kankan peer IDs can be used to uniquely identify different peers. *Extrapolating this value indicates that there should be well over 20 million unique users in a month for Kankan TV shows.* Also, it appears that the VoD component of the Xunlei network is more popular than the file downloading component. Nevertheless, Kankan swarms being larger than the BT swarms and over 7 million unique users in a week for just the TV show category certainly indicate a huge user base, probably at least as huge as the user base of BT.

When returning the list of Xunlei peers currently watching a queried Kankan movie, the Xunlei trackers return private IP addresses for peers coming from behind NATs. Out of the total 7,358,040 unique Kankan IDs collected, 44% of these peer IDs had private IP addresses. Assuming the Xunlei/Kankan users form a random sample of the Internet users in China, we can conclude that approximately 44% of Internet users are behind NATs.



(a) Snapshot of Swarm Size Distribution (b) Swarm Sizes of top 75 Xunlei and BitTorrent Movies

Fig. 4. Xunlei Swarm Scales

3.3 Geographic Distribution

We now examine the geographical distribution of Xunlei peers. We ran our DHT crawler for 72 hours and collected 5,381,746 peers with distinct \langle peer ID, IP, port \rangle . Those peers are from 140 countries, 3,935 cities and 2,052 ISPs.

Table 2 lists the top 15 countries with the most number of Xunlei peers. To no surprise, almost 98% of peers come from Mainland China. The population values are given in millions, and the last column shows the ratio of number of peers to the total population multiplied by 10^6 . Xunlei client being available only in Chinese language, this table provides some insight into relative Chinese populations in different countries. For example, Italy has a smaller population than UK and France, but yet has many more Xunlei users. This is consistent with recent reports of large Chinese immigration to Italy [1].

3.4 Content Flow from BitTorrent to Servers via Xunlei

Having observed a large fraction of BT content being tracked by Xunlei, we now investigate the flow of data from BT to Xunlei. To this end, we used our PirateBay Crawler to continually monitor the upload pages for two of the most active uploaders in Pirate Bay. As soon as a new TV show gets uploaded by either uploader, the BT infohash for the file is passed to our Xunlei Tracker Crawler, which then continuously crawls the Xunlei trackers to obtain the list of Xunlei peers and HTTP/FTP server resources every 5 minutes for several days.

None of 219 torrents we monitored were registered in the Xunlei tracker during the first crawl. However, 217 out of 219 were indexed by Xunlei trackers within 12 hours of the upload of the torrent to Pirate Bay, with 85% being indexed by Xunlei within less than 1 hour. We also observed that for 177 of the monitored torrents (80%), the Xunlei trackers returned at least one HTTP/FTP server within 6 days of appearance of the torrent in Pirate Bay. This means that the files corresponding to these torrents must have been uploaded to the servers. *In summary, for 177 of 219 popular torrents, the torrent first appeared in Pirate Bay, then in the Xunlei tracker without reference to a server, and finally in the Xunlei tracker with reference to one or more servers.* The following sequence of events can result in such a flow of data : 1) A Xunlei user obtains the .torrent file from Pirate Bay and then downloads the corresponding file using the Xunlei client. As a result of which, Xunlei trackers start tracking this particular file. 2) The same user (or some other Xunlei user who also performed the actions in step (1)) uploads the file to an HTTP/FTP server (e.g., to a cyberlocker, which often provides incentives to uploaders). This user advertises the link of the file in the server via an out-of-band channel, for example, a chat forum. 3) Another user finds the server link, and uses its Xunlei client to download the file. As a result, the server link is now listed as one of the resources for the file in Xunlei trackers.

Into what kinds of servers were these 177 files uploaded? The three domains serving the most files were megaupload.com, hotfile.com, and fileserve.com, serving 103, 96, and 61 files respectively. All three of these domains are cyberlockers. Although Xunlei can leech bandwidth from these (and other) cyberlockers, a user who directly visits a cyberlocker and uses Xunlei to download a file will quite possibly download parts from other sources, including other Xunlei peers. In this regard, Xunlei provides peer assistance to the cyberlockers.

4 The Gougou Content Index

We now analyze Xunlei from the perspective of content indexing site gougou.com. Using our Gougou Crawler, for each of 21,469 movies indexed by Gougou, we downloaded the corresponding Gougou page and extracted the meta data. Because each movie typically has many versions, we obtained meta data for 1,092,114 distinct files. Of these 21,469 movies, 34% came from Hollywood, 20% from China, 18% from Japan, 13% from Hong Kong, and the remaining 15% spread over several other countries.

We extracted the types of sources from each page of movies indexed by Gougou. HTTP (41%) and FTP (34%) sources clearly dominate the other types (MMS, RTSP, BT, ed2k, etc.) of sources available. BT takes the 3rd rank with about 10% of sources. For the HTTP and FTP sources, 24,513 and 11,642 different server IPs were found, respectively. The top HTTP server is sharing 23,727 files from three domains, and the top FTP server is serving 5,029 files under 73 different domains. A considerable fraction (12%) of FTP servers come from Chinese universities. A single FTP server located in a university in China is sharing as many as 2,701 files. Unlike Xunlei peers, a large fraction of sources originate from outside China, in particular 10% from the US.

5 Conclusion

We studied Xunlei, a download acceleration application that is enormously popular in China. We uncover the details of several unique technical features in Xunlei. By developing a comprehensive measurement infrastructure, we observed that the scale of content indexed by Xunlei is comparable, or perhaps, larger than the scale of content indexed by the most popular BitTorrent trackers. Similarly, we observed Xunlei/Kankan movie swarm scales to be comparable to that of BitTorrent. We found that HTTP and FTP sources make up the largest fraction of file sources indexed in Xunlei. Xunlei users benefit from increased content availability and increased download speeds. We also observed that Xunlei is currently providing bandwidth assistance to various cyberlockers and email distributions.

Acknowledgements

This work is partially supported by the NSFC under Grant No. 60972014, the fundamental research funds for the central universities under Grant No. HUST:2011QN015 and the technology support plan of China under Grant No. 2011BAK08B01. This work is also partially supported by NSF grant 0917767.

References

1. Chinese Remake the Made in Italy Fashion Label. <http://www.nytimes.com/2010/09/13/world/europe/13prato.html>.
2. Thunder Blasts uTorrent's Market Share Away. <http://torrentfreak.com/thunder-blasts-utorrents-market-share-away-091204/>.
3. P. Dhungel, K. W. Ross, M. Steiner, X. Hei, and Y. Tian. Measurement Study of Xunlei: Extended Version. Technical report, Polytechnic Institute of NYU, 2011. <http://cis.poly.edu/~prithula/papers/XunleiTR.pdf>.
4. M. Zhang, W. John, and C. Chen. A Measurement-Based Study of Xunlei. In *Student Workshop, PAM*, 2009.
5. M. Zhang, W. John, and C. Chen. Architecture and Download Behavior of Xunlei: A Measurement-Based Study. In *International Conference on Education Technology and Computer*, 2010.