Correlating topology and path characteristics of overlay networks and the internet

Citation for published version (APA):

DOI:
10.1109/CCGRID.2006.1630905

Document status and date:
Published: 01/01/2006

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Correlating Topology and Path Characteristics of Overlay Networks and the Internet
Alexandru Iosup, Paweł Garbacki, Johan Pouwelse, and Dick Epema
Delft University of Technology, Faculty of Electrical Engineering, Mathematics, and Computer Science
Mekelweg 4, 2628 CD, Delft, The Netherlands
Email: p2p@ewi.tudelft.nl

Abstract

Real-world IP applications such as Peer-to-Peer file-sharing are now able to benefit from network and location awareness. It is therefore crucial to understand the relation between underlay and overlay networks and to characterize the behavior of real users with regard to the Internet. For this purpose, we have designed and implemented MULTI-PROBE, a framework for large-scale P2P file-sharing measurements. Using this framework, we have performed measurements of BitTorrent, which is currently the P2P file sharing network with the largest amount of Internet traffic. We analyze and correlate these measurements to provide new insights into the topology, the connectivity, and the path characteristics of the Internet parts underlying P2P networks, as well as to present unique information on the BitTorrent throughput and connectivity.

1 Introduction

The topology and the characteristics of the Internet have a large impact on the operation of Internet applications such as multicast overlays [15], web hosting [8], and Peer-to-Peer (P2P) file-sharing systems [26]. In order to understand and improve the performance of these applications, it is essential to evaluate the way the underlay network supports the application overlays. In this work we present a measurement framework that allows joint large-scale measurements of the Internet and the BitTorrent P2P overlay, which is arguably the largest current Internet application. Large P2P networks continuously have more than 1,000,000 users1, and P2P file-sharing networks account for more than one third of the total Internet traffic [7, 12]. Over the past couple of years, BitTorrent has become the largest P2P file-sharing network in terms of generated traffic [12]. We therefore target our study on correlating the characteristics of BitTorrent and its Internet underlay. For this, we have designed and implemented MULTI-PROBE, a large-scale P2P measurement framework. With MULTI-PROBE, we build on the experience accumulated in our previous work [17, where we have investigated the high-level characteristics and the user behavior of BitTorrent, without correlating the underlying Internet. This work complements our previous investigations, adding the network level insights. For this, we focus on the following research topics:

- Measuring underlay/overlay networks
- Characterizing overlay networks and their users
- Correlating underlay/overlay measurements

1 Slyck.com lists size information for many large P2P networks.

Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid Workshops (CCGRIDW'06) 0-7695-2585-7/06 © 2006 IEEE
2 Background

In this section we provide a brief description of the BitTorrent P2P system, and we review related work.

2.1 A brief description of BitTorrent

BitTorrent is a P2P file-sharing network focusing on high data transfer speed rather than on search capabilities. BitTorrent is currently the largest P2P file-sharing network with over one third of the world’s P2P traffic [12], thus generating more than 15% of the total Internet traffic\(^2\). Furthermore, BitTorrent can efficiently handle flashcrowd effects [17] (offering popular data to a large number of users, immediately after the data release), and offers a pollution-free sharing environment to its users [17].

There are three levels in the BitTorrent world: the peer level, the tracker level, and the web-site level. Data, be it single files or file archives, exist in BitTorrent in the form of torrents. To facilitate the exchange process, torrents are split in smaller parts, called chunks. A user (peer) has a complete file only after obtaining all the chunks compositing that file. A swarm is the group of users downloading or uploading the same file; we define the swarm size as the number of users from a swarm, during a fixed time interval. BitTorrent peers are using an advanced tit-for-tat (bartering) mechanism to share file chunks. This mechanism ensures that for a user the amount of incoming data (bartering) mechanism to share file chunks. This mechanism ensures that for a user the amount of incoming data is roughly equal to the amount of outgoing data. The BitTorrent bartering-partners selection protocol aims at filling up all the available bandwidth of a peer; for popular files, the bandwidth filling property is guaranteed [19]. For this reason, and because of the lack of search features at the peer level, BitTorrent is bandwidth-limited. As opposed to other file-sharing protocols like DirectConnect and FastTrack, network latency is not an important issue for BitTorrent. By using the services of a tracker, BitTorrent peers can discover other peers that have chunks of the desired file. Since BitTorrent has no searching mechanism, metadata descriptions of the torrents are created in the form of .torrent files, which are accessible through web sites.

In time, BitTorrent has attracted a very specific, yet very large, audience. As a result of the excellent flashcrowd support, of the optimal pollution-removal system, of the existence of centralized services (the trackers), of a lack of search capabilities integrated in the peer functionality, BitTorrent has become a network for sharing fresh information. From the torrent files that can be found through the BitTorrent web sites, only the newest ones can be reliably and efficiently obtained from the network; the other files may be incomplete or even non-existent, or the number of peers that offer them can be too small to fill the desired download bandwidth. For non-fresh files, users typically resort to other networks, e.g., KaZaA or eDonkey, where locating files is easier, and old information can be more easily made available, at the expense of increased pollution [3].

2.2 Related work

The measurement-based analysis of Internet or large-scale P2P networks has been the subject of numerous investigations for over two decades.

In the area of Internet research, the CAIDA\(^3\) project is actively involved in periodic Internet mapping and characterization, and has arguably participated in the respective research community’s major achievements during the past 10 years. We mention two research results outside CAIDA: RouteViews and Rocketfuel. The RouteViews\(^4\) project collects BGP routing perspectives from more than 60 major ISPs worldwide. The combined table typically has nearly 120K globally routable prefixes. The Rocketfuel Internet mapping tool [23] uses traceroutes sourced at over 750 vantage points to explore the network topology of 10 ISPs in different continents. These efforts are not targeting any specific application; our BitTorrent study can offer additional insight into the actual use of the Internet, for instance by uncovering the peers’ download bandwidth (a comprehensive survey on the difficulties of accurately measuring the bandwidth available to and consumed by Internet applications is given in [18]), or by showing strong evidence that the P2P traffic of BitTorrent has shifted to non-default ports (answering the 2004 CAIDA question Is P2P dying or just hiding? [7] for the BitTorrent network).

Much effort has been put in characterizing protocols that optimize searching in P2P file-sharing networks, like Gnutella, FastTrack, or DirectConnect. Our study complements all these approaches by targeting a file download P2P protocol, BitTorrent. From the measurements targeting BitTorrent, our study is the first to characterize the peers sharing a representative set of files. In [21], the authors use active measurements to determine the bandwidth and latency of the peers of the largest P2P systems at that time, Napster and Gnutella. Their results only refer to P2P file-searching protocols. The passive-measurements of three popular P2P protocols, FastTrack, Gnutella, and DirectConnect, are analyzed in [22]. Again, results target only P2P file-searching protocols. In addition, their work suffers from the methodological aspect of monitoring only the default ports. In [6], a 5-month passive-measurements trace of a single file shared using the BitTorrent protocol is presented. The file comes from the operating system domain, and so is not representative for P2P file-sharing, where users download mostly movies and music.


\(^3\)The Cooperative Association for Internet Data Analysis (CAIDA). http://www.caida.org/.

A large-scale study of the eDonkey network is presented in [10]. The authors' focus is on clustering the peers' preferences; the peers' Internet connectivity is not analyzed, and from the peers' locations only the countries are detailed. Their analysis also displays a clear bias: the US-based users represent only 7% of the total world users, and users from Finland, Sweden, and China — countries strongly represented in terms of number of P2P users[6, 5] — represent each less than 1% of the total world users. In [16], the authors evaluate the traffic of ADSL users from an undisclosed location, and offer detailed statistics for a number of P2P networks. Their results are strongly influenced by the vantage point effect [13]. For example, they show evidence that the eDonkey network occupies more than 85% of the P2P traffic, which is impossible at world level, in the light of other accepted results [12, 21].

An attempt to correlate Internet and P2P network measurements is made in [9], where network connectivity and file-sharing between only 25 broadband users is studied. Therefore, that study lacks breadth, and should be complemented with a broader approach, such as this.

3 The MULTI PROBE framework

In this section we present the structure of MULTI PROBE, our framework for correlated measurements of large scale overlay networks and the Internet. MULTI PROBE consists of pre-processing, measurement, and post-processing parts, which we discuss below.

The BitTorrent network uses web sites as content directories. The largest of these sites handle tens of thousands of shared files and serve millions of users. MULTI PROBE can select the largest such web site, with respect to the number of users, at the start of the experiments (the SiteStats module, see Figure 1). Besides their directory service, the largest web sites also provide per-file information about the number of users uploading and downloading file chunks (swarm size). After sorting the shared files according to their swarm size, the measurements setup module, selects the top-N files from the largest web site, and their respective torrents are downloaded. The number of files measurement parameter (N) reflects the physical resources in the measurement infrastructure: the more measurement machines, the larger the possible value of N. In practice, the maximum value of N also depends on the number of files with swarms of sizes of a few tens of peers, since only these files can actually be reliably and efficiently obtained from the network [19].

For our measurements, we inserted monitoring nodes (probes) into the BitTorrent network. We define two types of measurements: active-start measurements, in which probes initiate the contact with BitTorrent peers, and passive-start measurements, in which probes wait for an externally initiated contact from BitTorrent peers. The need for the second approach is twofold. First, some measurement platforms limit the number of contacts a user can initiate during a fixed period of time. Second, the first approach does not allow to contact peers behind firewalls, while the second can overcome this limitation.

In active-start measurements, probes contact BitTorrent trackers repeatedly, acquiring lists of peer contacts (the GetPeers module, see Figure 1). They subsequently contact the corresponding peers using the BitTorrent protocol, and learn about what chunks are owned by these peers (the PingPeers module). In passive-start measurements, probes register themselves to a tracker, and wait to be contacted by other peers (the ListenPeers module). Discovery packets from multiple sources within the measurement infrastructure are sent to the peers discovered in this way, to characterize the Internet parts that are used by BitTorrent (the TrackPeers module); we call this process multi-source traceroute.

Finally, MULTI PROBE has a wide-range of post-processing modules, which can process and correlate detailed information about both the overlay network and the Internet characteristics.

4 Experimental Setup

In the pre-setup phase of the experiments, we first select the largest BitTorrent web site. We have developed size evaluators for four well-known BitTorrent web sites: Mininova (the replacement of the previously largest web site, SuprNova [17], which is now defunct), Novatina, Pirate Bay, and TorrentPortal. At the time of the measurements, these sites were the top four BitTorrent web sites in the slyck.com’s rankings. Table 1 shows the number of torrents (files) and users of each of these web sites; as Novatina was reporting exaggerated numbers of users and contained many broken links, we have removed it from further consideration. At the time of our measurements, the dominant BitTorrent web site was Pirates Bay. Therefore, for measuring topology and path characteristics of overlay networks and the Internet, we set up MULTI PROBE to measure the BitTorrent/Pirate Bay environment.

We have used two platforms for our measurements. For our active-start measurements, we have deployed MULTI-

<table>
<thead>
<tr>
<th>Web-site</th>
<th>Number of torrents</th>
<th>Number of users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pirates Bay</td>
<td>119,026</td>
<td>1,702,429</td>
</tr>
<tr>
<td>Mininova</td>
<td>21,790</td>
<td>1,014,332</td>
</tr>
<tr>
<td>TorrentPortal</td>
<td>26,580</td>
<td>848,704</td>
</tr>
</tbody>
</table>

Figure 1. The MultiProbe framework for correlated underlay/overlay measurements.

Table 2. A summary of the characteristics of the tracked files.

<table>
<thead>
<tr>
<th>Measurement type</th>
<th>Active-start</th>
<th></th>
<th>Passive-start</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg</td>
<td>Min</td>
<td>Max</td>
<td>Avg</td>
</tr>
<tr>
<td>BitTorrent swarm size</td>
<td>132</td>
<td>46</td>
<td>4,187</td>
<td>307</td>
</tr>
<tr>
<td>Torrent size</td>
<td>2.6GB</td>
<td>0.1MB</td>
<td>41GB</td>
<td>3.0GB</td>
</tr>
<tr>
<td>Number of torrent chunks</td>
<td>2,463</td>
<td>1</td>
<td>34,591</td>
<td>2,623</td>
</tr>
<tr>
<td>Torrent chunk size</td>
<td>1.1MB</td>
<td>32KB</td>
<td>4MB</td>
<td>1.2MB</td>
</tr>
<tr>
<td>Number of files in torrent</td>
<td>47</td>
<td>1</td>
<td>5,265</td>
<td>53</td>
</tr>
</tbody>
</table>

PROBE on 100 of the 400 processors available on our Distributed ASCI Supercomputer\(^6\), located in The Netherlands. For our passive-start measurements, we used 50 PlanetLab [14] nodes as peer points, and 300 PlanetLab nodes for traceroutes towards the peers that contacted the peer points. The 300 PlanetLab nodes were scattered around the world, but were mostly part of the GREN network [1]. Scriptroute [24] was of inestimable help for distributed software deployment and operation.

We successfully tracked the top 2,000 files in the active-start measurements. Due to machine failures on PlanetLab, we successfully tracked only 695 of the originally intended top 750 files in the passive-start measurements (files ranked 155, 331-340, 405-410, and 716-750 were not tracked). The two sets of tracked files overlap. Table 2 shows the statistics of the tracked files. We define a shared file as a file shared using BitTorrent. The set of tracked shared files presents a wide range of values for shared file size, number of chunks, chunk size, and files in the torrent. The average shared file size is around 2-3GB, which shows that BitTorrent is mostly used for large-sized transfers. We also find that most torrents are archives, and the average number of files per archive is around 50.

In this paragraph we quantify the bias introduced by our experimental infrastructure. We show that, unlike the previous approaches discussed in Section 2.2, our bias in negligible. First, the selection of the BitTorrent web site could induce a strong bias to the measurements, for instance if the selected web site is small-sized or serves only a special type of community, e.g., operating-systems developers. By selecting the largest BitTorrent web site, which covers a broad range of file types and sizes, we have ensured that we capture the characteristics of a world-wide community of users. However, as Pirates Bay is located in Sweden, we could not preclude the bias of the analysis of users’ locations towards Sweden. In Section 5.2 we quantify this bias, and we show evidence that it is negligible. Second, the selection of nodes located within the PlanetLab network could lead to biased results, as the PlanetLab nodes are much better connected than the average users of BitTorrent (see [1]). Because of this, there is a clear bias in round-trip time measurements; as BitTorrent is bandwidth-limited network, this bias is of minor importance. Furthermore, because of the BitTorrent’s bartering-partners selection protocol, connections tend to be formed with better-connected nodes. Indeed, a bias is introduced by the use of PlanetLab nodes; we argue that this bias is in fact required to ensure the representativeness of our results: we perform experiments from the points to which any peer in the BitTorrent network would naturally try to connect. This is key to a representative analysis of path structure, Internet Organizations distribution, and Autonomous Systems distribution (see Section 5.2). Also, the location of the DAS and the PlanetLab nodes does not influence the download bandwidth measurements (see also Section 5.3), nor the application-level connectivity measurements (see also Section 5.3). Third, the selection of the measurement parameter number of files \(N\) could influence our findings, particularly in the case of observed bandwidth. Since we have selected only files with swarms with
size over 40 (see Table 2), the reliability and efficiency of the data download are increased, and the average bandwidth is high. We argue that this is the real use of BitTorrent, and this is the context in which it should be evaluated.

5 Measurement Results

In this section we present the measurement results; all the results are available on the MULTIPROBE web site (see Section 8). For space reasons, we focus here on three main aspects. First, we correlate the users’ geographical locations with the generated traffic. The second category of results refers to route analysis and application-level connectivity; we report path characteristics of the overlay network flows, including path hops and AS mapping, and analyze application-level connectivity. The third and final category are related to traffic analysis. The break-down of traffic per TCP port shows that, contrary to reported traffic analysis results from 2004 [16], there is a wide amount of hidden traffic on non-standard ports. These results confirm the initial findings of [7], and corroborate them with quantitative data. We complete our traffic analysis with a detailed break-down of the application-level bandwidth.

5.1 Measurements summary

We record two types of events in active-start measurements: ContactOK events, observed when the probe successfully contacted a target peer, and ContactErr events, observed when a probe did not manage to contact a target peer. For ContactOK events, we record the target peer’s IP address, contact port number, and number of owned chunks. For ContactErr events, we record the target peer’s IP address, contact port number, and reason of the connection failure (e.g., connection timed out).

For passive-start measurement, there are two main types of events: Contacted events, observed when the measurement peer is contacted by another peer, and multi-source traceroute events, observed when the reverse path tree towards a peer was completed. We store detailed data relevant for these two types of events in sets of compressed files, one set per traced file.

Table 3 shows statistics of the size of our measurements. The two types of measurements yielded approximately the same number of identified unique users, over 225,000. Overall, we have gathered over 40,000,000 events. Note that the passive-start measurements, while hampered by the instability of the measurement machines, leads to the observation of almost as many unique users. We explain this phenomenon by the fact that the active probing technique requires more packets for the identification of each user.

5.2 Location analysis

In order to understand the links between overlay networks and the Internet, it is mandatory to geo-locate the users and to compare different locations according to the number of users and their traffic. We define a user’s weight as the amount of data transferred by that user.

We used the publicly-available libraries and databases provided MaxMind’s GeoIP7 and WebLogExpert’s8 to geo-locate the BitTorrent users. In our previous work [5] we have established the superior performance in both speed and IP address mapping accuracy of the combined use of these tools, when compared with two other similar approaches, CAIDA’s NetGeo9 and IP:Country; the reason is the more up-to-date information present in the GeoIP databases. From the original data, 3-4% of the IP addresses could not be mapped to continents and/or countries. As many as 21% of the IP addresses could not be mapped to a city, a known problem of the GeoIP (and NetGeo) database building approach [11]. We define an Internet organization as a corporate network or an ISP for home users. Less than 1.5% of the IP addresses could not be mapped to an Internet organization.

Table 4 shows the number of users and their weights for different continents and countries. The World rank row shows the relative country rankings. Europe is the dominant continent in both the number of users and the users’ weights. This result is not the result of the measurements bias; we have shown evidence of a similar phenomenon, but for a very limited number of files, in our previous work [5]. We expected Sweden to be placed in the top 10 countries, but its surprising ranking as 2nd in the world can only be explained as a bias of our data source (Pirates Bay is located in Sweden). In order to evaluate the bias, we resort to our previous results [5]: there, Sweden was occupying the 8th place in the world, with the percentage of users and weights per country of 2.6 and 3.3, respectively. Assuming these results as correct, the bias induced by the selection of Pirates Bay as the measurements target is +6.0% and +10.4%. Eliminating this bias, Europe remains (by far) the dominant continent, and Sweden remains in the world Top-10. We conclude that the bias was negligible.

Major cities like Madrid, Amsterdam, and Toronto are strongly represented in the top 50 cities both by number of users and by their weight. However, we found that buffer cities—small cities that host important network junctions, such as Oldenburg (DE), and Herndon (USA)—are also highly ranked, and cover 10% of the total users.

More than 6500 distinct Internet Organizations (IO) carry BitTorrent traffic. Figure 2 shows the CDF of the number of users and users’ weights per IO. As expected,

7MaxMind, http://www.maxmind.com
9http://www.caida.org/tools/utilities/netgeo/
Table 3. Statistics of the size of our measurements.

<table>
<thead>
<tr>
<th>Measurement type</th>
<th>Active-start</th>
<th>Passive-start</th>
<th>Measurement type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure</td>
<td>Value</td>
<td>Value</td>
<td>Measure</td>
</tr>
<tr>
<td>Period</td>
<td>05-11 May, 2005</td>
<td>Period</td>
<td></td>
</tr>
<tr>
<td>No. of files</td>
<td>2000</td>
<td>695</td>
<td>No. of files</td>
</tr>
<tr>
<td>No. of unique users</td>
<td>229,410</td>
<td>226,441</td>
<td>No. of unique users</td>
</tr>
<tr>
<td>Observed traffic</td>
<td>193.78 TB</td>
<td>909,961</td>
<td>No. of successful traceroutes</td>
</tr>
<tr>
<td>No. of ContactOK events</td>
<td>28,781,347</td>
<td>19,893,684</td>
<td>No. of visited IP addresses</td>
</tr>
<tr>
<td>No. of ContactErr events</td>
<td>7,240,601</td>
<td>3,729,570</td>
<td>No. of contacted events</td>
</tr>
<tr>
<td>Data size, compressed</td>
<td>3GB</td>
<td>8GB</td>
<td>Data size, compressed</td>
</tr>
</tbody>
</table>

Table 4. The percentages of users per continent and country.

<table>
<thead>
<tr>
<th>Continent, 2-letter</th>
<th>EU</th>
<th>NA</th>
<th>AS</th>
<th>SA</th>
<th>OC</th>
<th>AF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>59.4</td>
<td>22.3</td>
<td>8.0</td>
<td>6.9</td>
<td>2.7</td>
<td>0.5</td>
</tr>
<tr>
<td>Weight</td>
<td>58.8</td>
<td>22.1</td>
<td>7.4</td>
<td>4.5</td>
<td>2.6</td>
<td>0.3</td>
</tr>
<tr>
<td>Country (ISO3166)</td>
<td>SE</td>
<td>DE</td>
<td>ES</td>
<td>FR</td>
<td>FI</td>
<td>NL</td>
</tr>
<tr>
<td>Number</td>
<td>8.6</td>
<td>7.7</td>
<td>5.4</td>
<td>7.8</td>
<td>5.8</td>
<td>2.5</td>
</tr>
<tr>
<td>Weight</td>
<td>13.7</td>
<td>8.1</td>
<td>4.6</td>
<td>4.5</td>
<td>4.4</td>
<td>4.0</td>
</tr>
<tr>
<td>World rank</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Figure 2. The users’ Internet Organizations location: (a) CDF of number of users; (b) CDF of users’ weights. The horizontal axis shows the IO rank, with respect to their percentage.

There is no dominant IO; notably, the top 500 Internet organizations cover less than 90% in both categories. This shows that BitTorrent has grown to a large-scale, completely distributed state, in which both users and their traffic are scattered around the world.

Similarly to the IOs, more than 2400 different Autonomous Systems (AS) were identified as part of the BitTorrent’s underlay infrastructure; the top 250 ASs cover roughly 90% of the users. For space considerations, data regarding AS distribution is only available on the web site (see Section 8).

5.3 Route and connectivity analysis

Path IP hop count is a natural connectivity metric in the Internet [4]. Here, we also consider the number of AS traversals, as high latencies can be incurred at the interconnections between different ASs. Table 5 shows the distribution of IP path hop count, AS traversals, and intra-AS IP path hop count for more than 900,000 paths explored from PlanetLab peers to overlay network peers that have contacted them. The row Connection defines six categories of count ranges, with values given for IP path hop count, AS traversals, and intra-AS IP path hop count respectively in rows No. IP path hops, No. AS traversals, and No. intra-AS IP path hops. More than 66% of the users have a good connection to other peers or better, similar to the best hop count results reported in [4]. We also observe that a surprisingly high number of ASs have long routing paths inside: over 1% of the intra-AS paths have over 10 IP hops, due to faulty router configurations.

There is much interest in studying the round-trip time (RTT) characteristics of P2P networks. We measured the RTTs using traceroutes between our nodes and contacting peers. We report here only the results for successful traceroutes. Since measurements were conducted from well-connected PlanetLab nodes, there is a positive bias in the reported results: the average RTTs are smaller than normal. However, the fact that we used globally spread sources reduces this effect. Figure 3 shows the detailed distribution of RTT. More than 75% of the measured RTTs fall in the
Table 5. The distribution of IP path hop count, AS traversals, and intra-AS path hop count for paths between PlanetLab peers and overlay network peers which have contacted them.

<table>
<thead>
<tr>
<th>Connection</th>
<th>Direct</th>
<th>Strong</th>
<th>Good</th>
<th>Average</th>
<th>Loose</th>
<th>Very Loose</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. IP path hops</td>
<td>0-1</td>
<td>2-4</td>
<td>5-9</td>
<td>10-14</td>
<td>15-19</td>
<td>20+</td>
</tr>
<tr>
<td>No. paths [%]</td>
<td>5.9</td>
<td>20.4</td>
<td>40.2</td>
<td>25.3</td>
<td>6.4</td>
<td>1.1</td>
</tr>
<tr>
<td>No. AS traversals</td>
<td>0-1</td>
<td>2</td>
<td>3</td>
<td>4.5</td>
<td>6-9</td>
<td>10+</td>
</tr>
<tr>
<td>No. paths [%]</td>
<td>12.5</td>
<td>44.7</td>
<td>32.0</td>
<td>10.2</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>No. intra-AS IP path hops</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4.5</td>
<td>6-9</td>
<td>10+</td>
</tr>
<tr>
<td>No. paths [%]</td>
<td>58.1</td>
<td>12.9</td>
<td>8.0</td>
<td>12.2</td>
<td>7.6</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Figure 3. The distribution of measured RTT: (a) detailed distribution and zoom to RTTs below 1.5 seconds; (b) CDF of RTT. The vertical axis shows the respective percentages of numbers of RTTs.

100ms-1s category. Less than 10% of the measured RTTs are outside the 2.2 seconds margin. Since BitTorrent is not latency-limited, we conclude that most BitTorrent peers are close to each other, if the distance metric is RTT.

We define the application-level connectivity of a P2P network peer as the number of incoming and outgoing connections between that peer and other peers in the network. The maximum number of outgoing connections is a constant in most BitTorrent clients. We analyzed the number of incoming connections per 5 and 10 minutes, for all observed swarm sizes. For this, we first used the peer points from PlanetLab. According to the BitTorrent protocol, peers first contact a tracker, obtain a list of peers that are sharing the same file, then periodically initiate contact with a randomly selected peer from the list. The peer points are waiting for incoming connections and record them. We have then split the recorded data in 5 and 10 minutes intervals. Finally, we have correlated this information with the observed swarm size for the same intervals. We found that, regardless of the swarm size, for time intervals of 5 and 10 minutes the average number of incoming connections is around 10 and 20, respectively.

5.4 Traffic break-down

In order to monitor and control effectively P2P traffic, it is important to have a realistic traffic break-down per TCP port. Table 6 shows the distribution of users and user weights per TCP port. These results show that, contrary to reported distribution of users results from 2004 [16], there is a wide amount of hidden traffic on non-standard ports.

Figure 4. Average bandwidth per swarm size.

The high percentage of observed users of the standard BitTorrent ports (6881 through 6999) is misleading; the standard ports traffic accounts for a mere 11.2% of the total traffic. A number of non-standard ports (data not shown here) are used by high traffic users: ports 16881, 49152, 10000 and 10001 account alone for almost 2% of the total traffic. Given the overall distribution of traffic per TCP ports, we can conclude that BitTorrent clients have shifted in the last year to random port selection.

We define the application-level bandwidth as the total amount of data downloaded by a peer over a fixed time interval. Based on the BitTorrent’s bartering partners selection protocol (see Section 2.1), we argue that this measure is equivalent with the bottleneck bandwidth.

Figure 4 presents the influence of the swarm size on users’ bandwidth. The connected dots represent the aver-
Table 6. The distribution of users and user weights per TCP port.

<table>
<thead>
<tr>
<th>Start port</th>
<th>End port</th>
<th>No. users [%]</th>
<th>Weight [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>999</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>6881</td>
<td>6880</td>
<td>3.4</td>
<td>4.2</td>
</tr>
<tr>
<td>7000</td>
<td>6999</td>
<td>42.1</td>
<td>11.2</td>
</tr>
<tr>
<td>10000</td>
<td>9999</td>
<td>5.1</td>
<td>9.7</td>
</tr>
<tr>
<td>20000</td>
<td>19999</td>
<td>19.1</td>
<td>31.3</td>
</tr>
<tr>
<td>30000</td>
<td>29999</td>
<td>9.7</td>
<td>16.6</td>
</tr>
<tr>
<td>40000</td>
<td>39999</td>
<td>5.4</td>
<td>7.0</td>
</tr>
<tr>
<td>50000</td>
<td>49999</td>
<td>7.0</td>
<td>9.0</td>
</tr>
<tr>
<td>59999</td>
<td>60000</td>
<td>6.8</td>
<td>9.1</td>
</tr>
</tbody>
</table>

Figure 5. Application-level bandwidth: (a) number of users per bandwidth, value; (b) number of users per bandwidth, CDF.

In this section we discuss possible ways of using the data presented in this paper for modeling realistic P2P workloads, improving the way the Internet serves BitTorrent, and devising good measurement methods.

With the full release of our data (see Section 8), we hope that the research community dealing with Internet-aware P2P technologies [2, 20] will be able to test and/or improve their models with more realistic parameters. Our measurements address only the peer-level of a bandwidth-limited file-sharing network, but give insight in both the characteristics of the application (file types and sizes, bandwidth), and the way the application interacts with the underlay network (traffic size, paths, ports, even RTTs). The relatively small number of ASs and IOs needed to cover the majority of users (each less than 500) may enable detailed simulations even on small experimental setups. Other communities can use the information presented in this work: since BitTorrent has been shown to have better performance than the traditional file-transferring methods — in [25], BitTorrent and a vanilla flavor of it are shown to achieve much better scalability and transfer speed than (Grid) FTP, when used in an Internet-deployed (Desktop) Grid, and for files that exceed 20MB — we argue in favor of using our results for more realistic workload and environment modeling.

While the BitTorrent traffic is spread across the world, we find that less than 250 ASs and less than 500 IOs are used by the majority of users. In addition, over 25% of the intra-AS paths have 3-9 hops. Thus, making AS-level nodes BitTorrent-aware may lead to a reduction of intra-AS resource contention (routing), to an increase in the BitTorrent download performance (caching), and to advanced capabilities of traffic shaping (filtering).

BitTorrent is a bandwidth-dominated and bandwidth-aware application. Furthermore, BitTorrent connections are made to diverse port values. The implications of these properties on the design and conduction of experiments are twofold: the measurement infrastructure must be set such that it is well-connected to the Internet, and all the TCP flows must be observed, regardless of the source/destination port.

7 Conclusions and future work

This paper presents a correlated view of overlay networks and the Internet. For that purpose, we have designed, implemented, and deployed MULTI PROBE, a large-scale P2P measurement framework. Large-scale joint measure-
ments of BitTorrent and the Internet were conducted in May 2005, and correlated into comprehensive statistical data, in four categories: location, route, connectivity, and traffic.

Our main results are: (1) the majority of BitTorrent users are located in Europe but, at the same time, BitTorrent traffic is globally spread; (2) BitTorrent peers are well connected by the Internet in terms of path characteristics; (3) the average BitTorrent download bandwidth is 500 kbps; (4) the average BitTorrent application-level connectivity does not depend on the swarm size; (5) BitTorrent has shifted in the last year from static to random TCP port selection.

For the future, we plan to enable our framework to measure several other large-scale P2P networks, and to use MULTI PROBE for measurements in the new environments.

8 Availability

For space reasons, a number of graphs, tables and other results could not be presented in this paper. All this material as well as the data presented in this study can be found at the MULTI PROBE web site: http://multiprobe.ewi.tudelft.nl/.

Acknowledgments

Many thanks go to Neil Spring. We cannot stress enough how his excellent work on Scriptroute, and his ready help, allowed us to finish this project. The first author would also like to help Ana Lucia Varbanescu, who has been of utmost help.

References