Estimating P2P Traffic Volume at USC

ISI-TR-645, June 2007

Genevieve Bartlett† John Heidemann† Christos Papadopoulos‡ James Pepin†
†USC/Information Sciences Institute ‡Colorado State University
{bartlett,johnh,pepin}@isi.edu christos@cs.colostate.edu

Abstract

With the rise of peer-to-peer (P2P) file sharing applications there has been an increasing interest in understanding the popularity and use of P2P. In this study, we look at P2P use on the University of Southern California's campus network throughout a 14-hour period. We quantify the volume of traffic from P2P activity as well as the number of campus hosts involved in P2P at USC. Since port-matching techniques often fail for P2P applications, we estimate traffic based on both port-based and connection-pattern based techniques. We do not have access to packet data and so these measures provide only bounds on P2P traffic. In addition, while we identify P2P sharing, we cannot comment the types of data being shared (either music or data, restricted or freely available). We find that 3–13% of active hosts on campus participate in P2P, and that this traffic accounts for 21–33% of the bytes transferred to and from our campus.

1 Introduction

Since the emergence of peer-to-peer (P2P) file sharing applications in 1999 there has been a steady increase in the use of P2P file sharing. P2P sharing began primarily as a means to share music, often in violation of copyright laws. Today it is widely used to share a range of media, including music, video, and large data sets such as open source operating systems. It remains controversial, both because it supports a mix of illegal and legal content, and because widespread use can consume a significant amount of network bandwidth.

Although there has been a great deal of interest in understanding P2P file sharing, there has been relatively little quantitative measurement. A study at the University of Calgary is one of the few well documented studies, reporting an average of 38% of bytes of traffic represent P2P sharing [12]. While this study includes two years of data, it provided relatively little detail about who uses P2P and how many hosts were involved. Other reports are often anecdotal, leaving the details of the methodology in doubt, but reporting up to 80% of outgoing bytes at ISPs consisting of P2P traffic.

Following P2P trends has become important because P2P consumes significant network resources. In addition, there is great interest in understanding the degree of sharing of illegal content, and the opportunities and frequency of use of P2P to share legal content. Our goal in this report is to prove quantitative evaluation of P2P file sharing, including information about the relative use on academic-only networks like Internet2, and use by students relative to the general university population.

The following report discusses educated estimates of the amount of P2P file sharing detected over the University of Southern California's campus network. We monitor two (of two) commercial access links to Los Nettos and one (of two) Internet2 links. Of all traffic passing over these links, we analyze traffic for USC's two ranges of network addresses. Since P2P traffic often conceals itself in different ways we use two different, complementary methods to iden-

tify P2P traffic: first we identify P2P hosts based on communication over standard well-known P2P ports, second we use a novel technique that identifies P2P hosts based on their pattern of communication with other hosts, identifying patterns that are inherent to P2P activities. From this analysis of these packet traces we present approximate statistics about P2P traffic quantities at USC.

We estimate that 3–13% of active hosts at USC participate in P2P activities and account for *at most* 21–33% of the traffic volume at USC (Section 5.1). We also quantify activity on commercial networks as compared to academic networks like Internet2 (Section 5.3), and by some types of network access (wired, dormitory, etc., Section 5.4). We demonstrate that student lab networks and resident hall networks account for the majority of P2P activity at USC, indicating that students are the main users of P2P applications on campus. Because we do not have access to packet data, these measures provide only bounds on P2P traffic, we cannot comment the types of data being shared (either music or data, restricted or freely available).

2 Approaches to Identify P2P

Exact statistics on P2P traffic are difficult to obtain because detecting P2P activity is nontrivial. All detection schemes have limitations, however, by using two individual methods, we increase our ability to catch evasive P2P activity.

In this section we discuss our two methods of identifying P2P traffic: an improvement on using well known port numbers and a method based on inherent network behavior developed in our previous work [4]. In section 5 we discuss the P2P activity we found at USC using these two methods.

To limit privacy concerns, we use only *blind* techniques. These techniques rely on information in the TCP/IP headers in packets and do not look at any application level data.

The following two sections discuss two different approaches to identifying P2P hosts which use blind techniques.

Protocol	**Standard Ports**
BitTorrent	6969, 6881–6889
eDonkey	4661–4671
WinMx	6257, 6699
Gnutella	6346, 6347
Kazaa	1214

Table 1: Well known ports used by our port-based method.

2.1 Port Usage

The simplest method to identify hosts running a target application is to look at which port numbers a host has open for incoming connections. Firewalls and filters often rely on remote servers listening on one or more well known ports to determine which outgoing connections to block or allow. Port-based has in the past been an effective method to both identify hosts running a particular application and identify specific application flows.

While many well known protocols typically do listen on a well known port, such as port 80 for web traffic, applications which have reasons to hide often use non-standard ports to evade detection. It has become increasingly clear that it is not possible to identify P2P flows based solely on port numbers [6,9], and in fact port numbers are not useful in general for application detection [6].

With traditional well-known port number detection, we can easily identify P2P hosts which still listen on a standard P2P port. To enhance traditional port number detection, we also look for hiding peers to communicate with peers which still listen on a standard P2P port. With this enhancement, it only takes one connection to identify a host participating in P2P activities, even if the host is not listening on a standard P2P port.

The port numbers we use to identify P2P traffic are summarized in Table 1. We target five popular P2P protocols using 26 well known port numbers. We feel this set of ports is a complete set of well established popular P2P ports which non-P2P communication typically will not use.

2.2 Inherent Methods

Recent efforts to identify P2P traffic have produced several blind P2P detection methods which do not rely on standard port numbers [4, 7, 10, 11]. These methods focus on *inherent* network behaviors—behaviors which are not easily changed and are necessary behaviors for the application to achieve its goals. Our second blind technique is an inherent-network-behavior based method developed in our previous work [4].

Our inherent-network-behavior-based method focuses on three inherent network behaviors which many P2P applications share.

Failed Connections P2P peers contact a relatively large number of other hosts which do not respond. Since peers are end-user machines, there is considerable churn within a P2P network as peers come and go frequently [5]. Mechanisms which track the current membership of a peer group do so imperfectly, and as a result, peers often attempt to contact other peers which have already left the group.

Server and Client Behavior P2P hosts both make and accept TCP connections. Unlike typical client/server applications, such as web browsing applications, nearly all P2P applications have the ability to both make and accept connections. Typically, P2P peers attempt to quickly establish and keep a fixed number of incoming and outgoing connections to help maintain the interconnectivity of the peer network and avoid the peer network splitting into disjoint cliques.

Unprivileged Port Usage Often P2P peers communicate over connections which use unprivileged ports for both the source *and destination* of the connection. By convention, servers listen on standard privileged port numbers (below port number 1024), and clients make connections from unprivileged port numbers (above port number 1024). In contrast, P2P peers typically listen on ports above port number 1024 and also make connections on ports above port number 1024. A typical P2P peer will have a relatively high number of ongoing connections which use unprivileged ports for both source and destination.

Each of the three behaviors are exhibited in non-P2P protocols, but the combination of the behaviors is indicative of P2P activity. Furthermore, our method looks for the ratio of connections which fit the behavior to the total connections ongoing at a host. We look for hosts which have ratios within empirically derived thresholds, which helps reduce false positives. For example, a host which is scanning a network will exhibit failed connection behavior, but typically not at the same rate that a P2P peer will.

3 Implementation of Inherent Methods

In the previous section we presented three behaviors which are indicative of P2P activity at a host. In this section we present a brief overview of how we look for these behaviors in near real time. For full details, please see our paper [4] and the supporting ISI technical report [3].

We first define a metric for each behavior, which is then turned into a binary test used to confirm or disclaim P2P activity at a host. The binary test is positive for P2P activity if the ratio value for that behavior is within a lower and upper threshold, or negative if the ratio value is above the upper limit. If the ratio value is below the lower threshold, the test is inconclusive.

We consider connections over a sliding window of time. For each time window, we maintain a structure of host records containing an entry for each USC host which has new connection activity during the time window. As new connections are started during the time window, we update the record for the USC host.

To reduce the number of false positives, we wait until a minimum number of "warm-up" connections are made to and/or from a host before attempting to make a decision. Once a minimum number of connections are made, we test all three metrics in parallel. If all tests indicate positive, the host is flagged

as having P2P activity during that time window. If any of the tests indicate negative, the host is flagged as not running P2P during that time window. If no decision can be made, we continue to add new connections from that time window into the tests until the end of the time-window.

4 Data Collection and Evaluation Methodology

In this section we describe how we collected data to estimate the amount of P2P traffic at USC, while Section 5 presents the results.

Our evaluation uses USC network traffic captured from two of three commercial provider links at Los Nettos, a regional ISP, and one of two Internet2 links. Full network packet traces were collected during a 14 hour period from December 14th, 2006 at 9pm to December 15th at 11am.

Over the 14 hour monitoring period, we compile a list of all hosts detected via the port-based method as discussed in section 2.1 and all hosts detected via inherent methods discussed in section 2.2.

We also quantify the volume of P2P traffic at USC. Because both of our methods identify hosts participating in P2P and not individual P2P flows, our methods *can not directly calculate the P2P traffic volume*.

To estimate the P2P traffic volume, we count *all* bytes to and from an identified P2P host during the 14 hour monitoring period as P2P traffic. Counting all traffic to and from a host for the full monitoring duration is a conservative decision and will lead to an *overestimate of P2P traffic volume* because a host participating in P2P will likely also be running other applications.

5 P2P Activity at USC

In the following sections we present estimates on the amount and types of P2P activity at USC.

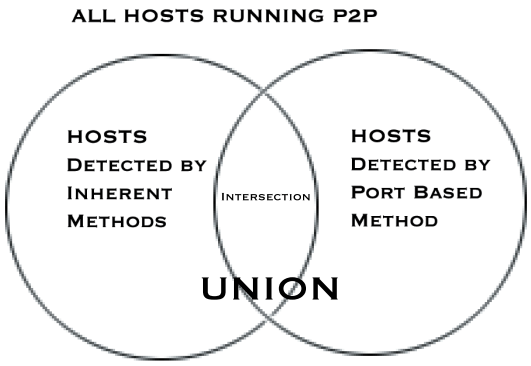

Figure 1: Venn diagram of P2P hosts

5.1 Estimating Total P2P Activity

We begin by establishing upper and lower estimates of the total amount of P2P activity on the USC network.

We first quantify how many hosts are detected in total (the union), by both methods (the intersection) and by only one of the two methods (see Figure 1). We claim the union represents an upper bound on the amount of P2P activity and the intersection represents a lower bound. We discuss the validity of this claim in the next section.

Table 2 summarizes the number of hosts and volume of traffic which was seen by our three monitoring points. The union shows an upper bound of 13% of USC's hosts participate in P2P activities and account for 33% of the total traffic volume seen over the three links we monitor. The intersection suggests a lower-bound estimate of only 3% of USC's hosts participate in P2P activities, accounting for 21% of the total traffic volume seen at our monitored links. From these results, it appears that prior reports that up to 80% of traffic is due to P2P applications do not apply to USC's university environment.

5.2 Comparison of Detection Methods

In the previous section we presented an estimate of the total amount of P2P activity present at USC.

	hosts	volume
Total	**16,120 (100%)**	**1,431 (100%) GB**
Identified as P2P (UNION)	2,051 (13%)	468 GB (33%)
Inherent-based Only	164 (1%)	26 GB (2%)
Port-based Only	1,423 (9%)	139 GB (10%)
Both (INTERSECTION)	464 (3%)	303 GB (21%)
Not identified AS P2P	14,069 (87%)	963 GB (67%)

Table 2: Summary of P2P activity at USC

We based our estimate on activity identified by two methods. In this section, we give perspective on the upper and lower estimates given in the previous section by comparing the P2P behaviors caught by each method.

We expect to see a significant overlap in the hosts detected by each method since both methods are designed specifically to detect P2P; however, we do not expect the overlap to be close to complete. Each method has limitations which cause the method to miss specific types of P2P behavior, and each method has separate causes for false identifications, causing a decrease in the intersection.

Table 3 summarizes the overlap between hosts identified by the two methods as participating in P2P activities and bytes identified as P2P.

Despite the limitations of each method, we claim that the union, with 2,051 hosts, represents a fair upper bound of P2P activity at USC. Each method's limitations in detecting P2P activity is offset by the other method. Our inherent-behavior-based method will capture active P2P hosts, even hosts which do not use standard or well-known P2P ports. Our port-based detection needs only one connection to detect a P2P host, and so is able to capture relatively idle P2P hosts which our inherent-behavior-based method may miss.

Because hosts in the intersection were identified by two separate and independent methods, we believe these 464 hosts represent a solid set of true positives, and offer a reasonable lower bound.

Port-based detection does not identify 26% of the 628 hosts identified by inherent-behavior-based detection. The fact that port-based detection misses a significant portion of hosts is not surprising since our port-based method will miss any P2P host which never communicates over a standard P2P port including hosts using P2P protocols which avoid using well established P2P ports all together.

Inherent behavior based detection does not identify 75% of the 1,887 hosts identified by port-based detection. It does, however, identify 68% of the 442 GB of traffic identified by port-based detection, implying the inherent based method catches the high-volume P2P hosts. Our inherent-based detection does miss idle hosts and is more sensitive to incomplete traffic views than our port-based detection. Our inherent-behavior-based method will miss P2P hosts which are relatively idle or which perform the majority of their P2P activity over an unmonitored link. (For the results presented in this paper, we require at least 10 new connections to be made at a host within a 20 minute time window in order to reach a decision.) However, missing idle peers does not greatly affect our estimates in the previous section since idle hosts do not contribute greatly to the P2P traffic volume.

We can quantify how many idle P2P hosts are missed by our inherent-behavior-based method by defining an idle P2P peer as any host which makes or receives relatively few connections (during our 14 hour monitoring period) over a well known P2P port. As seen in the CDF of the number of connections using a well known P2P port (Figure 5.2), of the hosts caught only by port-based detection, 67% made fewer than 20 connections over a known P2P port. Twenty connections is relatively few compared to other identified P2P hosts, and so the majority of hosts caught only by port-based detection can be considered idle peers. In contrast, of the hosts caught by both methods, only 18% had fewer than 20 P2P related connec-

Counting Method	Total Identified (Union)	Identified by Port-based	Identified by Inherent	Identified by both (Intersection)
Hosts	2,051	1,887	628	464
Bytes	468 GB	442 GB	329 GB	303 GB

Table 3: Summary of P2P Identified

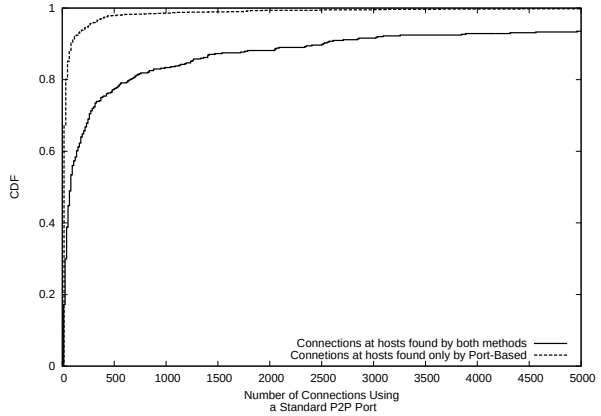

Figure 2: CDF of Number of connections using a standard P2P port

tions. We conclude that the majority of hosts missed by the inherent-behavior-based detection are missed because not enough actual P2P activity is captured.

5.3 P2P Activity on Commercial vs. Academic Networks

The previous two sections provided insight into the total amount of P2P activity at USC. In this section we give insight into who is participating in P2P networks by quantifying how much of the identified P2P activity is seen on the data collected with commercial ISP peers, versus how much is seen over one link with Internet2's Abilene Network.

We expect that the majority of P2P sharing is done between universities, and not between USC and commercial sites, for two main reasons: students are often attracted to P2P and university networks often offer high-bandwidth connections with light restrictions. The now defunct "i2hub" P2P service, which connected over 400 universities, is an example of the popularity of P2P sharing between university students [1].

Because the majority of Internet2's Abilene Network participants are universities, P2P activity on the monitored Internet2 link is likely traffic between USC and other universities. Therefore, we expect that the majority of P2P activity is over the Internet2 link.

Table 4 summarizes the traffic volume over the monitored Internet2 links and the two commercial provider links. Any traffic from a host identified as participating in P2P is counted as P2P traffic, as discussed in section 4.

The traffic monitored at the Internet2 link has a significantly higher percentage of P2P traffic than the commercial provider links (42% vs. 23%). This difference supports our claim that significantly more P2P traffic is inter-campus than between campuses and commercial sites.

5.4 Type of Host Participating in P2P

The previous section claims the majority of P2P traffic is over the Internet2 link because the majority of P2P users are students. To further prove this claim, in this section we look at which machines on campus are participating in P2P activities.

The subnetworks we monitored can be broken into six groups: wireless subnets, student lab subnets, student residence subnets, PPP subnets, VPN subnets and other subnets, which include internal operation machines.

When we look at the breakdown of how much P2P activity was found on the subnets in each group, we expect that a significant majority of P2P activity is found on the student related subnets, such as the residence subnets.

Link Type	Total	Ingress	Egress
Academic (Internet2 link)	**705 GB**	**658 GB**	**47 GB**
Identified as P2P (union)	299 GB (42%)	297 GB (45%)	2 GB (4%)
Inherent-based Only	16 GB (2%)	16 GB (2%)	0 GB (0%)
Port-based Only	103 GB (15%)	102 GB (16%)	1 GB (2%)
Both (intersection)	180 GB (25%)	179 GB (27%)	1 GB (2%)
Not identified AS P2P	406 GB (58%)	361 GB (55%)	45 GB (96%)
Commercial	**726 GB**	**70 GB**	**656 GB**
Identified as P2P (union)	169 GB (23%)	18 GB (26%)	151 GB (23%)
Inherent-based Only	10 GB (2%)	1 GB (2%)	9 GB (1%)
Port-based Only	36 GB (5%)	5 GB (7%)	31 GB (5%)
Both (intersection)	123 GB (16%)	12 GB (17%)	111 GB (17%)
Not identified AS P2P	557 GB (77%)	52 GB (74%)	505 GB (77%)

Table 4: Summary of traffic volume over monitored links.

Table 5 summarizes the break down of P2P activity by subnet groups. As expected, the majority of P2P traffic is detected on the student resident subnets, both by absolute volume and by percent of the total volume, with P2P accounting for 49–70% of resident hall traffic. By percent, the student labs also have a relatively high volume of P2P, with up to 45% of the total traffic identified as P2P. However, the overlap between the two methods for the student labs is nonexistent, implying that the P2P peers on the lab nets are idle.

These results again indicate that students are the main contributors to P2P activity in the university environment.

5.5 Ingress vs. Egress

The previous sections estimated the total P2P traffic volume at USC. In this section we look at the volume of P2P traffic leaving USC and the amount coming into USC to estimate to what extent USC provides content to P2P file sharing networks.

Due to less restrictions in a university environment, coupled with high bandwidth connections, we expect USC peers to be strong content providers for commercial hosts. Between universities, we expect that the sharing is more mutual.

The last two columns in Table 4 summarize the traffic entering and leaving USC over the Internet2 link and the commercial links.

The volume of traffic leaving USC over the commercial links implies that USC is generally a content provider to non-university hosts (of the 726GB seen over the commercial links, 656GB is traffic is leaving USC). The percentages of P2P traffic in either direction are roughly the same (26% of outgoing traffic is P2P, 23% of incoming is P2P), indicating that P2P data flow between USC and commercial sites is proportional to general data flow.

Over the Internet2 link, the incoming to outgoing ratio of P2P bytes is nearly 150GB to 1GB. This vast difference implies P2P sharing between USC and other universities is not mutual, with USC leeching more P2P content than it shares. However, this ratio could be skewed due to our monitoring view point.

5.6 Determining Popular P2P Protocols

The previous sections dealt with the amount of P2P activity at USC and the main contributors. In this section we look at which protocols appear to be popular and give insight into which protocols are easiest to detect.

We expect to see BitTorrent and Gnutella among the most popular applications. BitTorrent has a unique, and popular web-integrated system for directly connecting users interested in downloading

	hosts	traffic volume
Wireless (0–5% traffic due to P2P)	**1,024**	**0.039 GB (100%)**
identified AS P2P (UNION)	31	0.002 GB (5%)
Identified by Inherent-based only	0	0 GB (0%)
Identified by Port-based only	31	0.002 GB (5%)
Identified by both	0	0 GB (0%)
Not identified AS P2P	993	0.037 GB (95%)
Student Labs (0–45% traffic due to P2P)	**1280**	**40.346 GB (100%)**
identified AS P2P (UNION)	32	18.126 GB (45%)
Identified by Inherent-based only	2	0.051 GB (0%)
Identified by Port-based only	30	18.075 GB (45%)
Identified by both	0	0 GB (0%)
Not identified AS P2P	1248	22.220 GB (55%)
Residence Halls (49–70% traffic due to P2P)	**9,984**	**518.722 GB (100%)**
identified AS P2P (UNION)	1500	362.956 GB (70%)
Identified by Inherent-based only	467	21.628 GB (4%)
Identified by Port-based only	687	86.726 GB (17%)
Identified by both	346	254.602 GB (49%)
Not identified AS P2P	8484	155.766 GB (30%)
PPP (1–2% traffic due to P2P)	**1024**	**50.114 GB (100%)**
identified AS P2P (UNION)	53	0.906 GB (2%)
Identified by Inherent-based only	24	0.158 GB (0%)
Identified by Port-based only	17	0.332 GB (1%)
Identified by both	12	0.416 GB (1%)
Not identified AS P2P	971	49.208 GB (98%)
VPN (17–30% traffic due to P2P)	**1024**	**285.387 GB (100%)**
identified AS P2P (UNION)	402	85.643 GB (30%)
Identified by Inherent-based only	124	3.202 GB (1%)
Identified by Port-based only	174	34.172 GB (12%)
Identified by both	104	48.269 GB (17%)
Not identified AS P2P	622	199.744 GB (70%)
Other (0% traffic due to P2P)	**1,784**	**536.693 GB (100%)**
identified AS P2P (UNION)	33	1.238 GB (0%)
Identified by Inherent-based only	11	0.462 GB (0%)
Identified by Port-based only	20	0.473 GB (0%)
Identified by both	2	0.303 GB (0%)
Not identified AS P2P	1,751	535.455 GB (100%)

Table 5: Break down by subnet of identified P2P activity.

and/or sharing a specific resource [2]. Gnutella has a popular, long standing, network which connects millions of peers [13] through a tiered system.

Using our port-based method, we can estimate which protocols are used by the identified P2P hosts. Table 6 summarizes the protocol break down of the hosts identified by the port-based method, as well as the number of hosts which were also identified by our inherent-behavior-based method.

As expected, BitTorrent and Gnutella appear to be the most popular out of the five protocols our port-based method can identify. Though the overlap in host detection between the two methods is in the range of 0–13% for each of the protocols, the inherent-based method detects 79% of the bytes detected by the port-based method for the three most popular protocols (BitTorrent, Gnutella and eDonkey).

A large number of P2P peers used a mix of port numbers leading us to believe that a large number of P2P users do not have a single protocol preference and use multiple types of P2P applications.

There is also a greater overlap between the two methods for the multiple protocols category, indicating that there are fewer false identifications with either method when looking for hosts which use multiple P2P applications. This overlap is not surprising since a host is unlikely to contact or listen on multiple different well-known P2P ports unless it is involved in P2P activities. Also, if multiple P2P applications are run simultaneously, our inherent-behavior-based method is more likely to detect the host since, presumably, the P2P behavior is increased.

6 Related Work

In this section we briefly discuss other areas of research related to this work.

Closest to our work is a longitudinal comparison study done by Madhukar et al [12]. This study is a two year analysis of P2P activity at the University of Calgary and compares three methods of classifying P2P: a port-based method, a signature-based method and a blind technique based on work by Karagiannis et al [10]. Their findings suggest 30–70% of flows on their campus are due to P2P, with P2P responsible for an average of 38% of the bytes transfered. While our study does not compare methods of classification, our final findings offer a more in depth look at the users and sources of P2P in a university environment.

There are several bodies of work related to our inherent-based method which also look for inherent network behaviors such as communication patterns, protocol usage and failed connections [7, 8, 10, 11].

Similar to our port-based method, Wagner et al. look for hiding peers which are not listening on standard ports to contact non-hiding peers occasionally [14]. Their PeerTracker algorithm is successful at detecting the majority of high-volume P2P hosts.

7 Conclusion

Based on our 14 hour study, we estimate that 21–33% of USC's traffic is P2P related and 3–13% of the active hosts on campus participate in P2P activities. This estimate implies that USC participates in less P2P activity than other universities of comparable size.

Acknowledgments

We would like to thank Los Nettos for facilitating and granting access to trace collection. We would also like to thank Mark Baklarz and Steve Sutor of USC for discussions on current practices for tracking P2P users at USC.

This material is based on work supported by the United States Department of Homeland Security contract number NBCHC040137 ("LANDER"). It is also supported by the National Science Foundation (NSF) under grant number CNS-0626696, "NeTS-NBD: Maltraffic Analysis and Detection in Challenging and Aggregate Traffic (MADCAT)". All conclusions of this work are those of the authors and do not necessarily reflect the views of the sponsors.

References

[1] http://en.wikipedia.org/wiki/i2hub.

	All Identified (Union)	Inherent Method	Port-based Method	Both (Intersection)
Total	**2,051**	**628**	**1,880**	**464**
Protocol Estimated by Ports Used				
BitTorrent	341	44	341	44
Gnutella	866	145	866	145
Kazaa	55	3	55	3
eDonkey	210	28	210	28
Winmx	1	0	1	0
Multiple Protocols	414	244	414	244
Other	164	164	N/A	N/A

Table 6: Hosts identified as P2P broken down by protocol

[2] http://www.bittorrent.org/.

[3] G. Bartlett, J. Heidemann, and C. Papadopoulos. Inherent behaviors for on-line detection of peer-to-peer file sharing. Technical Report ISI-TR-627, ISI, 2006.

[4] Genevieve Bartlett, John Heidemann, and Christos Papadopoulos. Inherent behaviors for on-line detection of peer-to-peer file sharing. In *Proceedings of 10th IEEE Global Internet Symposium (GI '07) in conjunction with IEEE INFOCOM 2007*, Anchorage, AK, USA, May 2007.

[5] M. Bawa, H. Deshpande, and H. Garcia-Molina. Transience of peers and streaming media. In *Proceedings of the ACM HotNets I*, pages 107–112, Princeton, NJ, USA, October 2002.

[6] K. Cho, K. Fukuda, H. Esaki, and A. Kato. The impact and implications of the growth in residential user-to-user traffic. In *Proceedings of the ACM SIGCOMM Conference*, pages 207–218, Pisa, Italy, September 2006.

[7] M. Collins and M. Reiter. Finding peer-to-peer file-sharing using coarse network behaviors. In *Proceedings of the European Symposium On Research In Computer Security*, Hamburg, Germany, September 2006.

[8] F. Constantinou and P. Mavrommatis. Identifying known and unknown peer-to-peer traffic. In *IEEE International Symposium on Network Computing and Applications (NCA)*, pages 93–102, Cambridge, MA, USA, July 2006.

[9] T. Karagiannis, A. Broido, N. Brownlee, kc claffy, and M. Faloutsos. Is P2P dying or just hiding? In *IEEE Global Internet*, 2004.

[10] T. Karagiannis, A. Broido, M. Faloutsos, and kc claffy. Transport layer identification of p2p traffic. In *Proceedings of the ACM SIGCOMM Workshop on Internet Measurement (IMC)*, pages 121–134, Taormina, Sicily, Italy, October 2004.

[11] T. Karagiannis, K. Papagiannaki, and M. Faloutsos. BLINC: Multilevel traffic classification in the dark. In *Proceedings of the ACM SIGCOMM Conference*, pages 229–240, Philadelphia, PA, USA, August 2005.

[12] A. Madhukar and C. Williamson. A longitudinal study of P2P traffic classification. In *Proceedings of the International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, pages 179–188, September 2006.

[13] A. Rasti, D. Stutzbach, and R. Rejaie. On the long-term evolution of the two-tier gnutella overlay. In *IEEE Global Internet*, 2006.

[14] A. Wagner, T. Dubendorfer, L. Hammerle, and B. Plattner. Flow-based identification of p2p heavy-hitters. In *Proceedings of the ICISP International Conference on Internet Surveillance and Protection*, August 2006.