On The Identification and Analysis of P2P Traffic Aggregation*

Trang Dinh Dang, Marcell Perényi, András Gefferth, and Sándor Molnár

High Speed Networks Laboratory, Dept. of Telecommunications & Media Informatics
Budapest University of Technology & Economics
H-1117, Magyar tudósok krt. 2, Budapest, Hungary,
Email(s): {trang,perenyim,geffertha,molnar}@tmit.bme.hu

Abstract. The main purpose of this paper is twofold. First, we propose a novel identification method to reveal P2P traffic from traffic aggregation. Our method is based on a set of heuristics derived from the robust properties of P2P traffic. We show the high accuracy of the proposed algorithm based on a validation study. Second, several results of a comprehensive traffic analysis, focusing on the differences between P2P and non-P2P traffic, are reported in the paper. Our results show that the unique properties of P2P application traffic seem to fade away during aggregation and characteristics of the traffic will be similar to that of other non-P2P traffic aggregation.

1 Introduction

From the beginning of the new millennium the Internet traffic characteristics show a dramatic change due to the emerging *Peer-to-Peer* (P2P) applications. Starting from the first popular one (Napster) a number of new P2P based multimedia file sharing systems have been developed (FastTrack, eDonkey, Gnutella, Direct Connect, etc.). The traffic generated by these P2P applications consumes the biggest portion of bandwidth in campus networks, overtaking the traffic share of the world wide web [24, 2]. A common feature in all of these P2P applications is that they are built on the P2P system design where instead of using the server and client concept of the web each peer can function both as a server and a client to the other nodes of the network. This principle involves the adapting nature of P2P systems as individual peers join or leave the network. Another common feature of these P2P systems is that they are mainly used for multimedia file sharing (movies, music files, etc.), which frequently contain very large files (megabytes, gigabytes) in contrast to the typical small size of web pages (kilobytes).

A number of studies have been published in the field of P2P networking. Papers [1–4] focus on the measurement of different P2P systems like Napster, Gnutella, KaZaA, and the traffic characterization and analysis of P2P traffic

* The research was supported by Ericsson Traffic Lab and Magyar Telekom Ltd., Hungary.

providing some interesting results of resource characteristics, user behavior, and network performance. Several analytic efforts to model the operation and performance of P2P systems have been presented so far. Queueing models are applied in [5, 6], while in [7–9] branching processes and Markov models are used to describe P2P systems in the early transient and steady state. P2P analysis using game theory is presented in [19] among others. Other studies, e.g. [10–13], are concerned with the effective performance and the QoS issues of P2P systems. In addition, many papers [14–18] indicate various possible applications using P2P principles. Further approaches propose structured P2P systems using Distributed Hash Table (DHT) with several implementations like Pastry, Tapestry, CAN, Chord [20].

The P2P traffic characteristics are not fully explored today and there is a tendency that they will be even more difficult to analyze. In contrast to the first generation P2P systems the recent popular P2P applications disguise their generated traffic resulting in the problematic issue of *traffic identification*. The accurate P2P traffic identification is indispensable in traffic blocking, controlling, measurement and analysis. However, the issue is touched upon in only a few papers and the proposed solutions still have some drawbacks. The problem is that P2P communications are continuously changing, from TCP layers using well-known ports in some first versions to both TCP/UDP with arbitrary and/or jumping ports nowadays. A robust and accurate P2P traffic identification is vital for network operators and researchers but today there is a lack of published results on this field and this is our main motivation for the work presented in this paper.

The workload characteristics of peers participating in some P2P systems has been examined in several papers as mentioned before. However, from the aspect of service providers only little useful information can be gained from these studies. The service providers are less interested in the detailed activities of some particular P2P softwares but the traffic generated by peer users. This paper, in contrast, concentrates on those factors and characteristics of P2P communications which have an impact on the P2P traffic aggregation.

The rest of the paper is organized as follows. We describe our measurements and the pre-processed data in Section 2. Section 3 presents our heuristic P2P identification method. The traffic characterization results are given in Section 4. Finally, Section 5 concludes the paper.

2 Traffic Measurements

The measurements were taken at one of the largest Internet providers in Hungary in May 2005. In the chosen network segment, traffic of ADSL subscribers is multiplexed in some DSLAMs before entering the ATM access network. Placed at the border of the access network and the core network are some Cisco routers. NetFlow measurements are carried out at two of these routers in three days from May 26th to 28th. NetFlow, developed by Cisco, collects all *incoming* flow information and exports the logs periodically. The obtained data traces are the aggregate incoming traffic of more than 1000 ADSL subscribers.

Data sets	05_26	05_27	05_28
Time of measurement	05/26 17h15 - 05/27 7h	05/27 17h06 - 24h	05/28 0h - 24h
Number of flows	4 293 394	5 858 756	17 224 625
Total traffic [GB]	113.91	126.06	316.91

Table 1. Summary of the collected data sets

Three data sets were selected for analysis, which are denoted by 05_26, 05_27, and 05_28. The summary of the data sets is presented in Table 1.

3 P2P Traffic Identification

A number of published papers have dealt with the issue of P2P traffic identification. Port-based analysis is presented in [25]. [22, 26] provide a method using the relationships between P2P flows or P2P client/server. Identification based on application signatures is shown in [21, 23]. In addition, [23] also proposes another method of identification based on some heuristics. In summary, P2P traffic identification has two promising approaches:

- P2P traffic identification based on payload information
- P2P traffic identification based on flow dynamics

The first method can provide very high detection accuracy in case of well-known open P2P protocols. It takes advantage in the investigation of some named P2P systems. Its drawbacks appear in high processor claim (for payload check), and the continuous change of P2P protocols, which are not available in most of the cases. Moreover, it also raises a number of legal and privacy problems. The second one is simpler to perform but it implies heuristic methods yielding less accurate results. However, it does not depend directly on actual P2P systems, thus it is more consistent and suitable for the analysis of P2P traffic aggregation. In this paper we have chosen the second approach and present an accurate and robust simple P2P traffic identification method.

3.1 A heuristic method for P2P traffic identification

Our proposed heuristic method consists of six steps, each being associated with a group of P2P flows to be identified. At the beginning we try to classify a set of widely used Internet applications (except P2Ps) based on well-known port analysis.

Initial step While port based analysis is less accurate to identify P2P traffic, it is still appropriate to distinguish traffic of common applications. Our exhausted search of these applications and their communication ports, in both TCP and UDP layers, results in a table of application ports. Flows with these ports in the *source_port* or *dest_port* are first extracted from the data sets. HTTP/SHTTP ports are not among these. The reason is that HTTP ports are not only used

for web surfing but also by some P2P applications, e.g. KaZaA. The separation of web and P2P traffic is considered by the second heuristics.

Step 1 The first heuristics is based on the fact that many P2P protocols, e.g. eDonkey, Gnutella, Fasttrack, etc., use both TCP and UDP transport layers for communication. Reasonably the unreliable UDP is often used for control messaging, queries, and responses while data transmission relies on TCP. However, the large volume of UDP traffic observed in our measurement data indicates that UDP could also be used for data transfer. Thus by identifying those IP pairs which participate in concurrent TCP and UDP connections we can state that the traffic between these IP pairs is almost surely P2P.

This heuristics is similar to what is proposed in [23] with a little difference. We note that some other common applications like NETBIOS, DNS also utilize both TCP and UDP. [23] needs a post-processing to extract this kind of traffic from the result of the heuristics. In contrast, this is not necessary in our case since we have already done this in the initial step: these applications are among the common ones.

Step 2 The second heuristics tries to separate web and P2P traffic from flows using HTTP/SHTTP ports, i.e. 80, 8080, 443, etc. The typical difference between P2P and web communication of two hosts can be observed. In general, web servers use multiple parallel connections to hosts in order to transfer web pages text and images (also music, video contents in some cases). In contrast, data transmission between peers consists of one or more consecutive connections, i.e. only a single connection can be active at a time. This property is used to identify web servers and then the traffic originating from them.

The traffic using HTTP ports is divided into groups of individual IP pairs. The web server is the one with the IP address in the HTTP ports side which has parallel connections to its pair. We also differentiate between two cases: if the IP address of the web server belongs to the outside IP domain it is likely to be a public web server. Then all the HTTP traffic from them is marked as web traffic. In the other case only parallel flows with HTTP ports are marked as web traffic. The rest of this traffic group is P2P traffic.

Step 3 In the next step, P2P traffic is selected using default ports of P2P applications. P2P software often defines default ports for communication. It is true that in most cases peer users can change it to any arbitrary port (but it is not frequent since peer-to-peering is usually not prohibited for home users) or port can be dynamically chosen automatically or when firewall or port-blocking is observed. This step cannot detect all P2P connections, but once the traffic is collected we can be almost sure that it is from those concerned P2P systems.

A table of well-known ports used by some popular P2P applications is collected for this step. Flows containing these values in *source_port* or *dest_port* are all marked P2P.

Step 4 In normal TCP/UDP operation, at least one of the two ports is selected arbitrarily. It is not likely that flows with similar flow identities (*source_IP, dest_IP, source_port, dest_port, prot_byte, tos*) exist in relatively short measurements. This happens, however, in the case of P2P connections, if both source

and destination peers dedicate a fixed port for data transfer. File download of a file is often executed in several smaller chunks. Therefore multiple flows with the same flow identities can be generated by P2P software. This is the basis of this heuristics: the identical flows are from P2P applications if at least two of each are found.

Step 5 For the same reason as the above heuristics, it is not probable that a host (IP) will repeatedly choose a given arbitrary port for TCP/UDP connections unless it is a server. Web servers and other common server traffic is extracted by the previous heuristics, thus it is safe to introduce the next heuristics: if an IP uses a TCP/UDP port more than 5 times in the measurement period that {IP,port} pair indicates P2P traffic. The selected upper threshold (5) is a rule of thumb established empirically.

Step 6 The last heuristics is based on the fact that objects of P2P downloads often have large sizes from several MB in case of music files or smaller applications to hundreds of MB in case of video files and larger software packages. In addition, peer users are patient. P2P downloadings can last some ten minutes or hours. By this heuristics those flows are considered P2P flows which have flow size larger than 1 MB or flow length is longer than 10 minutes.

4 Analysis Studies

In this section, we first verify the robustness of the proposed P2P traffic identification method. Next the results of the identification of the measured data sets are presented. The detailed comparison analysis of P2P and non-P2P traffic aggregation follows.

4.1 Verification of the identification method

In order to examine the robustness of the heuristics presented in Section 3.1 a validation measurement was carried out. In this measurement besides gathering general and aggregated information of the traffic flows we also recorded the name of the corresponding application. This enabled us to validate the correctness of the proposed P2P traffic identification method.

The measurement collected the traffic generated by two Linux PCs running SMTP and web servers among others (although with very light traffic), and some P2P applications: *qtorrent*, *valknut*, and *aMule*. These are the Linux versions of the Bittorrent, Direct Connect and eDonkey systems, respectively. To challenge the identification method, we used non-default P2P ports. Several downloads were initiated, while the P2P clients were also enable to serve requests of other peers. The measured trace contains more than 120000 data flows.

We present the performance of each heuristics and the overall identification process in Fig. 1. The hit rate of each heuristics, counted in percentage, is the ratio of the number of *correctly marked* flows and the total number of *marked* flows by the heuristics. We note that the hit rate of the 6th heuristics is not shown in the figure because it marked no flows in this data set. The last two

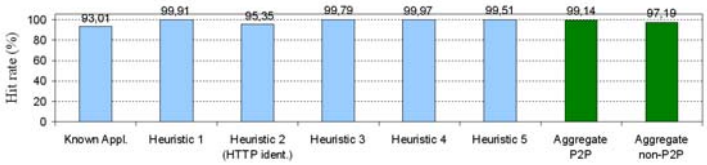


Fig. 1. Validation result of the identification method

columns in the figure show the rate of correctly marked P2P (non-P2P) flows and the total number of P2P (non-P2P) flows in the data set. The result is very convincing in every statistics. The average hit rate is greater than 99.7%. The amount of unidentified traffic is about 0.1%. The ratio of wrongly marked P2P flows and unidentified P2P flows per the total marked P2P flows are 0.3% and 0.8%, respectively. Note that these performance parameters are counted flow-wise. Similar results concerning the traffic quantities (bytes) are much better.

4.2 Traffic identification

As described earlier the traces are the sets of flow information collected using Cisco NetFlow measurements (see Section 2). We assign a flag to each flow record of our database. The flag has the default value of u which means unknown (traffic) and it can be changed in the course of the identification process. The list of possible values of the flag is the following:

u: default value, unchanged if the flow cannot be classified
m: management flow (classified by IP addresses of the routers)
o: other non-TCP/UDP flow (ICMP, IPv6, RSVP, etc.)
k, kh: known common application (except HTTPs), flow using HTTP ports
pX: P2P flow, X denotes the heuristics which identifies the flow

The result of the identification procedure is summarized in Table 2.

Data sets	05_26		05_27		05_28	
Flag	#flows [%]	volume [%]	#flows [%]	volume [%]	#flows [%]	volume [%]
m	0.5	0.01	0.08	0.005	0.1	0.007
o	0.42	0.06	0.87	0.05	0.88	0.29
k, kh	64.75	52.61	33.66	23.7	30.65	32.13
pX	33.82	47.29	64.94	76.19	68.05	67.51
u	0.5	0.03	0.43	0.05	0.03	0.06

Table 2. Traffic identification result

4.3 Traffic analysis

In this study the analysis framework focuses on the fundamental differences between the P2P traffic and other Internet traffic (this will be referred to as

non-P2P traffic). The comparison is done regarding several aspects of the traffic characterization.

Overview of the traffic The daily fluctuation of the traffic is presented in Fig. 2. The upper plot shows the total and non-P2P traffic intensities of the 05_28 data set while the lower one is the intensity and the flow count of the P2P traffic of the same set.

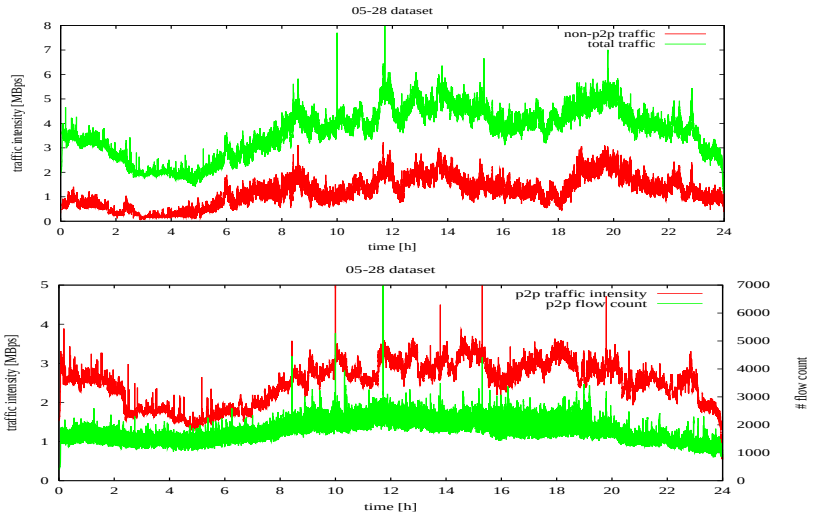


Fig. 2. Traffic intensities from 05_28 data set

As observed in general, daily traffic can be divided into two parts: the busy period from around 8h to 24h and the non-busy period from about 0h to 8h. Both P2P and non-P2P traffic follow this daily tendency. However, in the case of non-P2P applications the traffic level shift between busy and non-busy periods is significant (the bandwidth falls to very low values in non-busy period) while this ratio for P2P applications is only around 1/3. This is reasonable since non-P2P users, in general, do not generate traffic in the sleeping time. In contrast, P2P users (in our case also home users) turn on the P2P application and request some audio and video files (some can be very large). Then they leave the system to work over days, even when they are asleep during the night period. Basically, the P2P traffic can be steady over time, which can be seen in Fig. 2: the number of P2P flows has small variation (see the lower plot). We still see a certain decrease in the traffic. It happens since the number of downloadable sources decrease and probably more requests are not added during the night period.

The volume of P2P traffic, see also Table 2, which is about 65% of the total traffic, exceeds by far the traffic volume of the non-P2P applications.

Number of P2P and total active users In the measurement environment, Internet subscribes do not have fixed IP addresses. Each time a user connects to the Internet, a dynamic address is given to the user. Therefore it is impossible to determine exactly which data flow belongs to which user. However, less error is expected when we choose to associate an individual IP address to a user. Since the ADSL contracts at the present Internet provider do not limit the time of connections, the average connection time is relatively long. We assume that during our measurements, which lasted at most 24h, only a minimum number of IP address wanderings occurred.

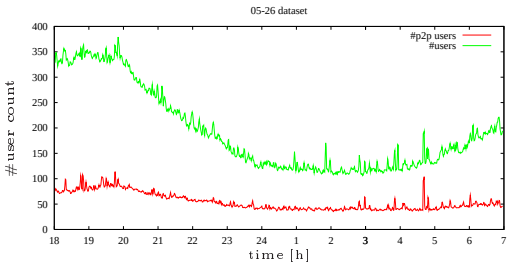


Fig. 3. The average number of (P2P) users (05_26 data set)

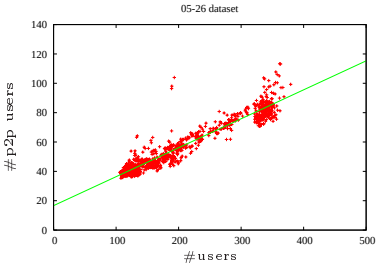


Fig. 4. Relation between P2P users and the total user number (05_26 data set)

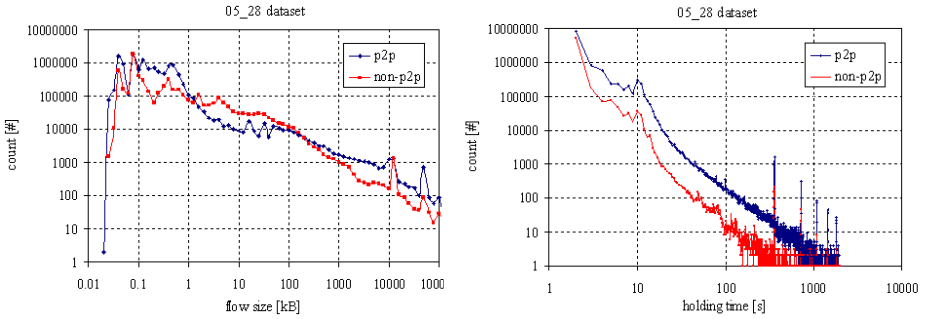
To calculate the number of active users, the number of different IP addresses participating in the flows is counted in every second. Then a sliding window of size 120s and step 50s is applied to smooth the variations caused by communication breaks. One of the results is shown in Fig. 3. The total number of users, according to the time shift between busy and non-busy periods, decays as the non-busy period is approached. The lowest number of users is observed in the non-busy period. This similarity is not so striking in the case of P2P users. The answer is similar to the above, it is due to the typical behavior of P2P users/applications.

The relation between the active P2P users and the total active users is presented in Fig. 4. As seen in the figure, 05_26 data set for example, there is a strong linear connection between the two measures. This means that approximately a fixed ratio of active users is using P2P applications. This is quite an interesting finding and it is hard to find a reasonable explanation. However, if this relation is general, it would be very useful for e.g. traffic dimensioning. We plan to verify this relation in more different network environments. The estimated ratio between P2P users and total users is about 0.2 for this data set, 0.3 for the other two sets.

The relation between the number of active (P2P) users and the occupied bandwidth is also investigated. It is shown that a linear connection can be observed in both cases (P2P and non-P2P traffic). However, the variance of data

around the assumed linear function is much higher than in the previous case. In addition, variation is higher and the slope of the line is much lower for non-P2P traffic. The same number of non-P2P users occupy much lower bandwidth compared to that of P2P users.

Flow sizes and holding times The next comparison is about the properties of data transferring: flow size and flow holding time. Fig. 5(a) presents the histogram of the flow sizes of P2P and non-P2P applications. We find no significant divergence in this characteristics. In both cases the plots, disregarding flow sizes smaller than 0.1 kB, nearly follow a straight line in the log-log scale. This indicates a possible heavy-tailed (Pareto) model for the flow size for both P2P (with shape parameter a=-0.3) and non-P2P flows (a=-0.25) and also for the overall traffic. (The assumptions of Pareto distribution were verified by several heavy-tailed tests: De Haan's moment method, Hill estimator, and QQ-plot [27].) The number of P2P flows which are larger than about 100 kB is somewhat higher than the number of non-P2P ones, which is also reasonable, but the difference is not significant.



(a) (b)

Fig. 5. Histogram of flow sizes (a) and flow holding times (b) (05_28 data set)

The result seems to be reconcilable with some newer developments of many P2P protocols. Independently of the size of the requested objects, at the beginning the P2P application downloads only a small chunk of the object. The condition of the network and source capacity is predicted from the characteristics of the previous downloads. The size of the next chunk will be determined according to the assumed download quality. Thus, at the end, the P2P traffic (concerning flow size in this case) behaves similarly as the non-P2P traffic.

Similarity is also obtained in the flow holding time distribution of P2P and non-P2P traffic, see Fig. 5(b). Again, in the log-log scale, one can see two almost parallel lines in the two histograms. The plots suggest the Pareto distribution for both cases with the same shape parameter a=1.4. The shift in the histogram

plot agrees with the fact that the total number of P2P flows is higher than that of the non-P2P ones by one order of magnitude.

Popularity distribution The IP addresses were ranked according to their total amount of downloaded traffic. The downloaded traffic is plotted against the ranked IP address (which we have assumed to be associated with an individual user) in Fig. 6. The skewness in the popularity distribution of P2P systems is also justified in our analysis as in many studies of P2P traffic. The top 10% of P2P users corresponds to more than 90% of total download traffic. Our interest, however, is how it differs from the other Internet traffic. Our analysis shows that the difference does not lie at the head of the rank but at the tail. As we go down the rank, the download traffic by ranked users decreases very fast in the case of P2P users. There is a big split between "obsessive" and hobby P2P users. In contrast, the degree of traffic volume decay in case of ranked non-P2P users is very slow. The average non-P2P users create relatively stable traffic when they access the Internet: reading daily news, chatting with friends, etc.

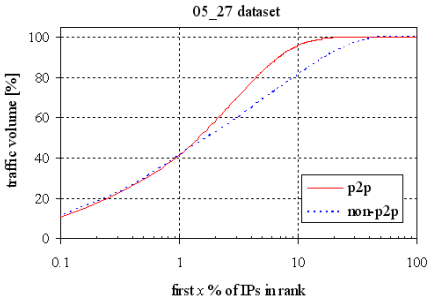


Fig. 6. Traffic volume of ranked IPs (05_27 data set)

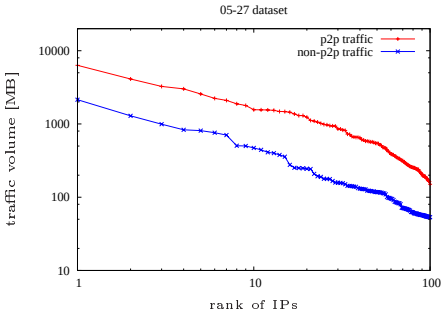


Fig. 7. Traffic vs. top ranked IPs of 05_27 data set

At the top (about 10%) of the ranked list the popular Zipf's law seems to be accurate to describe both P2P and non-P2P traffic popularity. As seen in Fig. 7 two almost linear plot of P2P (marked by +, upper curve) and non-P2P IP rank (marked by x, lower curve) with an approximate slope of -1 indicates the standard Zipf distribution as the suitable model for *top ranked* users' traffic.

Analysis was also carried out for the connection population and similar curves were shown in the results. Fast decrease was observed in the case of P2P traffic as the ranking place increases, the decay is much lower in non-P2P case. In average a normal non-P2P user creates more, and probably smaller connections than P2P users despite that P2P traffic dominates in all measurements both in the volume and the connection number. This happens because, for example, opening of a web page involves multiple downloads of text, many images, and even audio and video elements.

5 Conclusion

In this paper we first presented a novel P2P traffic identification method. The method collects a set of rules derived from the general behavior of P2P traffic. Our method does not use any payload information so it is easy to implement and use when payload cannot be evaluated because of legal or privacy obstacles or cannot be measured due to technical or financial problems. Our validation results show that the proposed algorithm is able to capture the P2P traffic very efficiently. The identification method was used to identify P2P traffic in current measurement data taken from one of the largest Internet providers in Hungary.

We also presented a comprehensive traffic analysis of P2P and non-P2P traffic. The obtained results have highlighted some critical findings. P2P users/applications, by the typical content-sharing objectives of P2P usage, behave in a different way than other Internet applications. The difference manifests itself in the almost stable P2P activities over busy and non-busy time periods, the bandwidth-hungry nature, the skewness in the traffic volume distribution between P2P users, etc. However, the characteristics of P2P traffic aggregation, which would be a more important aspect from the service providers' and network operators' point of view, are quite similar to those of other traffic aggregation. While in the beginning P2P applications were confined to greedy file-sharing, nowadays they have grown up to be an unisolable component of the Internet due to several refined developments of P2P protocols. It has been shown that there is always a certain ratio of home users who use some P2P applications. The study establishes that the workload of P2P applications generates similar (heavy-tailed) flow size and flow holding time distribution like several non-P2P applications. As a consequence the P2P aggregation also shows a similar characteristics.

There may come the time when we should change the way of thinking about and treating P2P traffic. It is not an outstanding but an inseparable part of the overall Internet traffic just like every other traffic component. Our future work will focus on the research to further investigate this conjecture for general Internet traffic.

References

1. S. Saroiu, K. P. Gummadi, R. Dunn, S. D. Gribble, H. M. Levy, "An Analysis of Internet Content Delivery Systems", in Proc. 5th Symposium on Operating Systems Design and Implementation, Boston, MA, USA, Dec. 2002.
2. S. Sen, J. Wang, "Analyzing Peer-to-Peer Traffic Across Large Networks", *IEEE/ACM Transactions on Networking*, 12(2):219-232, 2004.
3. K. Tutschku, "A Measurement-based Traffic Profile of the eDonkey Filesharing Service", PAM 2004: 12-21.
4. J.A. Pouwelse, P. Garbacki, D.H.J. Epema, H.J. Sips, "The Bittorrent P2p File-Sharing System: Measurements And Analysis", 4th Int. workshop on Peer-to-Peer Systems (IPTPS'05), Feb. 2005.
5. Z. Ge, D. R. Figueiredo, S. Jaiswal, J. Kurose, D. Towsley, "Modeling Peer-Peer File Sharing Systems", in Proc. INFOCOM'03, San Francisco, CA, Mar. 2003.

6. K. K. Ramachandran, B. Sikdar, "An Analytic Framework for Modeling Peer to Peer Networks", in Proc. INFOCOM'05, 2005.
7. G. de Veciana, X. Yang, "Fairness, Incentives and Performance in Peer-to-Peer Networks", in Proc. Allerton Conf. on Communication, Control and Computing, 2003.
8. X. Yang, G. de Veciana, "Service Capacity of Peer to Peer Networks", in Proc. INFOCOM'04, 2004.
9. D. Qiu, R. Srikant, "Modeling and Performance Analysis of BitTorrent-Like Peer-to-Peer Networks", in Proc. ACM SIGCOMM'04, Portland, OR, Aug. 2004.
10. B. Yang, S. Kamvar, H. Garcia-Molina, "Addressing the Non-Cooperation Problem in Competitive P2P Systems", Workshop on Peer-to-Peer and Economics, Jun. 2003.
11. D. Hughes, I. Warren, G. Coulson, "Improving QoS for Peer-to-Peer Applications through Adaptation", in Proc. of the 10th Int. Workshop on Future Trends in Distributed Computing Systems (FTDCS 2004), Suzhou, China, May 26-28, 2004.
12. E. Kalyvianaki, I. Pratt, "Building Adaptive Peer-To-Peer Systems", in Proc. 4th Int. Conf. on Peer-to-Peer Computing (P2P'04), 2004.
13. M. Iguchi, M. Terada, K. Fujimura, "Managing Resource and Servent Reputation in P2P Networks", in Proc. 37th Annual Hawaii Int. Conf. on System Sciences (HICSS'04), 2004.
14. G. Ding, B. Bhargava, "Peer-to-Peer File-Sharing over Mobile Ad hoc Networks", in Proc. PerCom Workshops, 2004.
15. M. Demirbas, H. Ferhatosmanoglu, "Peer-to-Peer Spatial Queries in Sensor Networks", in 3rd IEEE Int. Conf. on Peer-to-Peer Computing (P2P'03), Linkoping, Sweden, Sept. 2003.
16. M. Roussopoulos, M. Baker, D. S. H. Rosenthal, T. J. Giuli, P. Maniatis, J. C. Mogul, "2 P2P or Not 2 P2P?", IPTPS 2004: 33-43.
17. Y. Guo, K. Suh, J. Kurose, D. Towsley, "A Peer-to-Peer On-Demand Streaming Service and Its Performance Evaluation", in Proc. IEEE Int. Conf. on Multimedia & Expo (ICME 2003), Baltimore, MD, Jul. 2003.
18. G. Cugola, G. P. Picco, "Peer-to-Peer for Collaborative Applications", Int. Workshop on Mobile Teamwork Support, Vienna, Austria, Jul. 2002.
19. C. Buragohain, D. Agrawal, S. Suri, "A Game Theoretic Framework for Incentives in P2P Systems", in Proc. 3rd Int. Conf. on Peer-to-Peer Computing, 2003.
20. T. Risse, P. Knezevic, A. Wombacher, "P2P Evolution: From File-sharing to Decentralized Workflows", *Information Technology*, 4:193–199, Oldenbourg, 2004.
21. S. Sen, O. Spatscheck, D. Wang, "Accurate, Scalable In-Network Identification of P2P Traffic Using Application Signatures", in Proc. 13th Int. Conf. on World Wide Web, NY, USA, 2004.
22. M. Kim, H. Kang, J. W. Hong, "Towards Peer-to-Peer Traffic Analysis Using Flows", DSOM 2003: 55-67.
23. T. Karagiannis, A. Broido, M. Faloutsos, K. Claffy, "Transport Layer Identification of P2P Traffic", in Proc. 4th ACM SIGCOMM Conf. on Internet Measurement, Taormina, Sicily, Italy, Oct. 25-27, 2004.
24. Internet2 NetFlow: Weekly Reports - http://netflow.internet2.edu/weekly/
25. A. Gerber, J. Houle, H. Nguyen, M. Roughan, S. Sen, "P2P The Gorilla in the Cable", in National Cable & Telecommunications Association (NCTA) 2003 National Show, Chicago, IL, June 8-11, 2003.
26. S. Ohzahata, Y. Hagiwara, M. Terada, K. Kawashima, "A Traffic Identification Method and Evaluations for a Pure P2P Application", Lecture Notes in Computer Science, p55 Vol. 3431, 2005.
27. S. I. Resnick, "Heavy Tail Modeling and Teletraffic Data", *The Annals of Statistics*, 25(5):1805–1869, 1997.