Characterization of BitTorrent Traffic in a Broadband Access Network

Zoltán Móczár and Sándor Molnár

High Speed Networks Laboratory
Dept. of Telecommunications and Media Informatics
Budapest Univ. of Technology and Economics
H–1117, Magyar tudósok krt. 2., Budapest, Hungary
moczar.zoltan@gmail.com, molnar@tmit.bme.hu

Abstract. BitTorrent as one of the leading P2P file sharing applications has dominant traffic in broadband access networks. In this paper we present the main characteristics of BitTorrent traffic based on actual measurements taken from a commercial network. Analysis results at both application- and flow-levels are presented and discussed.

Key words: traffic measurements, traffic analysis, BitTorrent

1 Introduction

BitTorrent [1] is one of the most popular peer-to-peer (P2P) file sharing application in the Internet. It was designed by Bram Cohen in April 2001 and the first implementation was released on July 2, 2001. After a few years, BitTorrent reached extremely high popularity with 10 million users in 2005, and we can identify that the popularity increases continuously in the previous five years.

The mechanism of BitTorrent is based on the organization of peers sharing the same file in a P2P network, and this method ensures an efficient replication to distribute the shared file. The file is divided into little chunks, and a peer can perform a simultaneous download of more chunks. The exchange of file chunks is motivated by an incentive mechanism, which enables peers with high uploading bandwidth to reach high downloading bandwidth. Such mechanism can prevents free riding effectively, which is very common in P2P systems.

A majority of studies focuses on the identification of P2P traffic including the popular P2P applications like BitTorrent, Gnutella, and eDonkey [2, 3, 4]. There are several other studies addressing the characteristics of BitTorrent traffic. Choffnes and Bustamante pointed out that testbed-based views of Internet paths are surprisingly incomplete concerning BitTorrent and many other applications. This message gives us a warning and emphasizes the need for using actual measurements for analysis [5]. In the last years, several articles have been published, which analyze the behavior of BitTorrent. For example, Erman et al. presented a study on modeling and evaluation of the session characteristics of BitTorrent traffic [6]. They found that session inter-arrivals can be accurately modeled by the hyper-exponential distribution while session durations and sizes can be reasonably well-modeled by the log-normal distribution. Andrade et al. studied three BitTorrent content sharing communities regarding resource demand and supply. The study introduced an accurate model for the rate of peer arrivals over time. They also found that a small set of users contributes most of the resources in the communities, but the set of heavy contributors changes frequently and it is typically responsible only for a few resources used in the distribution of an individual file [7]. Authors in used full packet traces collected at a large edge network. They focused on three flow-level metrics, namely flow size, flow inter-arrival time and flow duration. In addition they also studied three host-level metrics: flow concurrency, transfer volume and geographic distance. Based on the results they presented flow-level distribution models.

The main goals of the paper are towards a better understanding of the characteristics of BitTorrent traffic found in broadband access networks at both application- and flow-levels. We have carried out a comprehensive analysis study based on measurements taken from a commercial network.

Our results shows the dominance of the BitTorrent traffic and we also found that the BitTorrent mechanism with the optical high bandwidth symmetrical access produced a double amount of upload BitTorrent traffic volume compared to the download traffic volume. The flow level results shows a detailed characterization concerning BitTorrent flow size, flow duration and number of packets including both TCP and UDP flow components of BitTorrent traffic.

The paper is organized as follows. In Section 2 we discuss the details of measurements including the network architecture and analysis tools. Section 3 and Section 4 present our analysis results at application- and flow-levels, respectively. Finally, Section 5 concludes the paper with our main results.

2 Traffic Measurements

Our measurements were taken from one of the commercial networks in Stockholm, Sweden. This company maintains network infrastructure for several service providers. These ISPs provide many different services like Internet access, IP telephony or IPTV for both residential and business users. The traffic in the measurements was generated by more than 1800 customers.

The network infrastructure of Swedish backbone network and the related residential network are shown in Fig. 1. The network company did not give any further information regarding subscribers due to privacy issues.

The backbone network consists of three core routers linked to each other with 3 Gbps optical fibres. The subscribers are connected to the area switches and their traffic is aggregated in a migration switch through 100 Mbps links. The migration switch is linked to one of the core routers with a 1 Gbps capacity link.

The workstation responsible for data capturing was connected to one of the core routers with a 1 Gbps capacity fibre. The router mirrored its traffic to the workstation, which let the capturing device store and dump the data on its hard drives. Only the packet headers were captured to get information for the analysis such as protocol, size and direction.

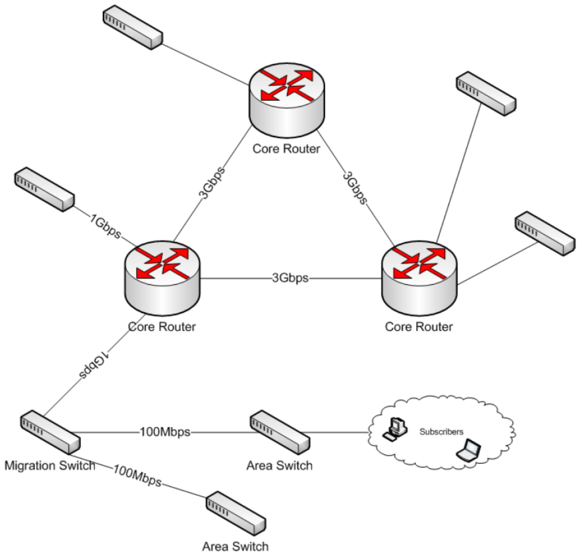


Fig. 1. The architecture of the measured network

Traffic identification was done with a tool developed by Ericsson Hungary. This software uses various techniques to identify the traffic such as port-based, signature-based and heuristic-based approaches, however, the algorithm is not public.

Table 1 describes the main parameters of the investigated traces where FL denotes the flow-level data sources. After the preprocessing phase the cleaned data were loaded into database tables using Microsoft SQL Server 2005. Data retrieving was performed by SQL queries, and the results were processed by Matlab routines. Moreover, Matlab was also used for visualizing and creating charts.

Table 1. Basic description of measurements

Trace	**Measurement period** October, 2008 [(day) hour:min]	**Duration** [hour:min]	**Flows** [million]	**Packets** [million]
FL-1	(7) 11:18 – (8) 22:16	34:59	59.1	3892
FL-2	(7) 15:00 – (7) 15:59	01:00	1.53	69.96

3 Application-level Analysis of BitTorrent

In this section we give some basic description about the measured traffic including the daily profile and user ranking based on the number of incoming and outgoing packets followed by a detailed discussion on the main characteristics of BitTorrent.

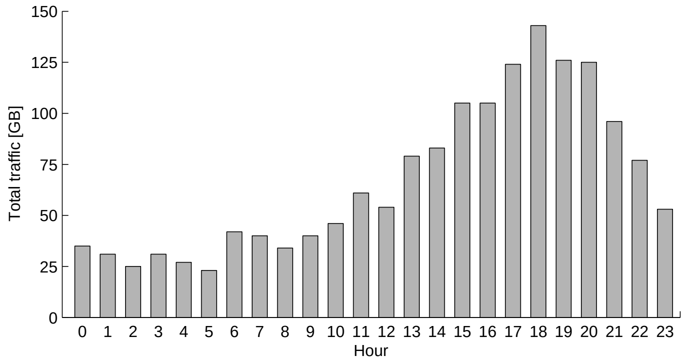


Fig. 2. Daily profile (FL-1)

The daily profile of the traffic is given in Fig. 2. We can see that the peak hours are between 17 PM and 20 PM. In this time frame the traffic volume is about six times higher than in non-busy hours in the mornings around 5 AM.

Fig. 3 shows the measured incoming (downlink) packets as a function of the measured outgoing (uplink) packets. Every point in the figure corresponds to one user. The linear fitting tells that an average user generates approximately 25% more downlink packets, though the variance is quite large.

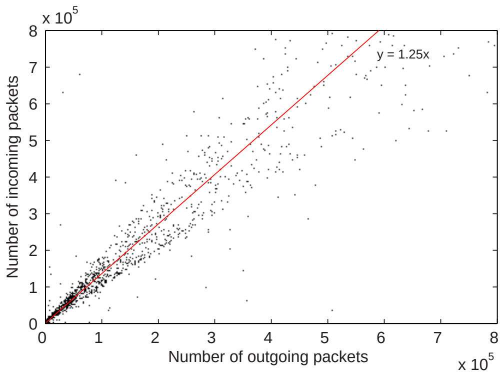


Fig. 3. User ranking based on the number of incoming and outgoing packets (FL-1)

Table 2 represents the incoming, outgoing and total traffic volumes generated by BitTorrent. It clearly indicates that the amount of data uploaded by BitTorrent users is more than the downlink traffic volume. More precisely, 68% of the total BitTorrent traffic is uplink traffic and this is two times higher amount of data compared to the BitTorrent downlink traffic. This is due to the usage of BitTorrent since the finished downloads stay in the queue and they are automatically shared with other users. However, the users can remove these shared files from the queue, but fortunately most of them do not do that what is crucial for the efficient operation of the BitTorrent network. Consequently, their uplink traffic increases by the finished and shared downloads. The volume of uplink traffic is not limited since most of the users have symmetric optical high bandwidth access.

Table 2. The traffic volume generated by BitTorrent (FL-2)

Incoming	Outgoing	Total
7.5 GB (32%)	15.8 GB (68%)	23.3 GB

4 Flow-level Analysis of BitTorrent

In the following analysis both incoming and outgoing traffic have been considered and results are related to the total traffic. In the investigated one and a half day long period 1217 GB was downloaded (incoming traffic) and 1568 GB was uploaded (outgoing traffic). We observed that the dominance of BitTorrent uploading determines the general picture and makes the uplink traffic volume 30% higher than the downlink traffic volume. Additionally, it gives the explanation of why the ratio is different for the number of incoming and outgoing packets (see Fig. 3), which is in contrast to the previous results.

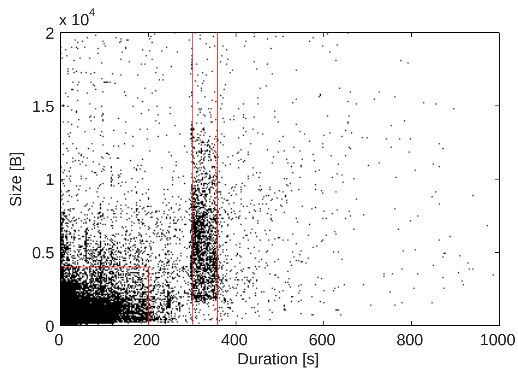


Fig. 4. Relationship between flow size and duration (FL-2)

Fig. 4 illustrates the relationship between flow size and duration for BitTorrent in an enlarged view. Every point in Fig. 4 and Fig. 5 represents exactly one flow. There are two different clusters in Fig. 4: a cluster concentrated around the horizontal and vertical axis, and another one bounded by the vertical lines. Our analysis showed that the bytes carried by a BitTorrent flow are almost independent of the flow duration. We found that in the first cluster 98.3% of BitTorrent flows have a duration less than 200 s and a size smaller than 4 kB. Furthermore, 85.2% of these flows are transferred over UDP and only 14.8% of them are sent over TCP. Concerning duration the flows of the second cluster fall in the interval 300 s and 365 s. In this region about 97.2% of flows are related to TCP and only a negligible amount of data is transmitted over UDP.

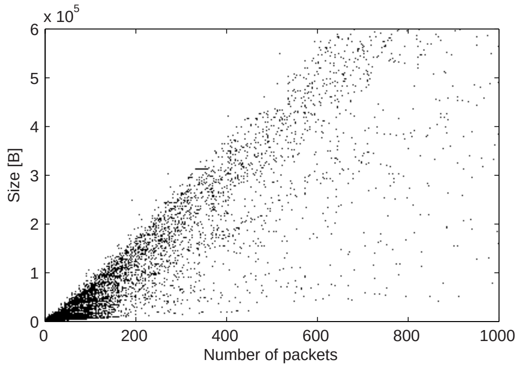


Fig. 5. Relationship between flow size and number of packets (FL-2)

Fig. 5 depicts the relationship between flow size and number of packets. The size of the largest packet is near to the maximum size of the Ethernet frame close to 1500 bytes, and it can be observed that BitTorrent flows are highly scattered in the examined dimensions.

Fig. 6 shows the histogram of the number of packets. We can observe that the number of BitTorrent flows have a heavy-tailed decrease for the number of packets more than 100. More deeply analysis revealed the interesting property that for the whole measurement period only 0.1% of flows contain a unique number of packets.

Fig. 7 shows the relationship between flow size and duration for TCP and UDP flows, respectively. BitTorrent is the most dominant P2P application in FL-2 contributing 46.8% of the total traffic. We can observe a peculiar difference between TCP and UDP flows in this figure. In case of TCP flows the graph is composed of two clusters: a cluster mainly consisting of horizontal lines and filling up almost completely the region between size 90 B and 6 kB and duration 100 μs and 320 s, and another cluster covering a large range in size between 1.5 kB and 580 MB, but a small range in duration between 5 s and 50 min. In contrast to TCP we can only find a single cluster in case of UDP flows, which is

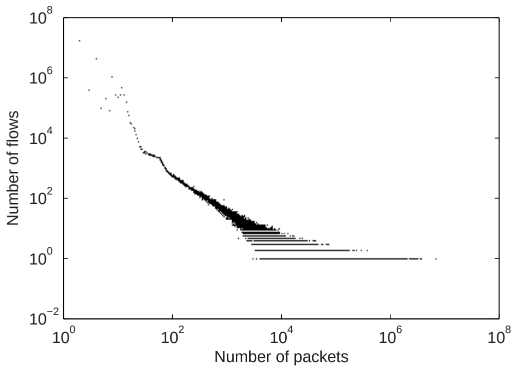
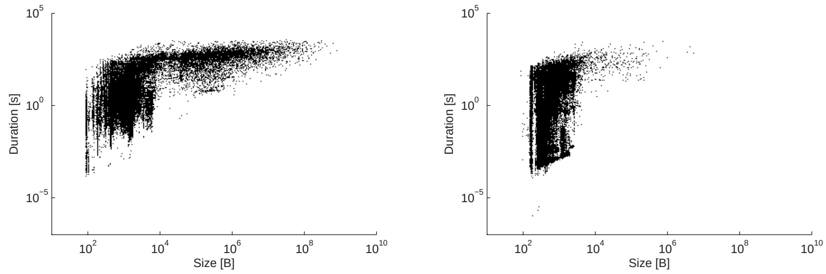


Fig. 6. Histogram of the number of packets (FL-1)



(a) TCP flows

(b) UDP flows

Fig. 7. Relationship between flow size and duration (FL-2)

similar to the related cluster concentrated around the horizontal lines in Fig. 7a. For the whole measurement period we got that although, only 16% of BitTorrent flows are transferred over TCP and 84% of them are sent over UDP, TCP flows carry almost 99% of the total bytes. The reason is that BitTorrent basically uses TCP for file transfer, but some of the new clients implement a UDP-based method to communicate with the tracker servers. Obviously, it does not need to transmit large volume of data, but rather needs to send numerous flows.

5 Conclusion

In this paper we presented a comprehensive traffic characterization study of BitTorrent based on actual measurements taken from a commercial network.

We found that the total number of incoming (download) packets is 25% higher than the total number of outgoing (upload) packets. However, the traffic volume shows a different picture: the outgoing traffic volume is about 30% higher than the incoming traffic volume. This observation is explained by the dominance of BitTorrent usage where the amount of data uploaded by an average user is much more higher than the downlink traffic volume in many cases and it was well supported by the symmetric optical high bandwidth access of most of the users. It resulted in double amount of upload traffic volume of BitTorrent compared to the download traffic volume.

The flow-level analysis revealed that the bytes carried by a BitTorrent flow are almost independent of the flow duration, and almost all of the flows have a duration less than 200 s and a size smaller than 4 kB. We also found that almost all of the flows, which have a duration between 300 s and 365 s are TCP flows. The size of the largest packet is near to the maximum size of the Ethernet frame close to 1500 bytes, and BitTorrent flows are highly scattered in the dimensions of size and number of packets. Furthermore, the number of BitTorrent flows have a heavy-tailed decrease for the number of packets more than a certain value, and only 0.1% of flows contain a unique number of packets. Although, only 16% of BitTorrent flows are transferred over TCP and 84% of them are sent over UDP, TCP flows carry almost 99% of the total bytes.

Acknowledgement

We thank Sollentuna Energi AB for the measurements, Ericsson Sweden and Ericsson Hungary for the cooperation and helping to access the data. The research was supported by NKTH-OTKA grant CNK77802.

References

1. BitTorrent homepage, http://www.bittorrent.org/index.html
2. Sen, S., Spatscheck, O., Wang, D.: Accurate, Scalable In-Network Identification of P2P Traffic. In: WWW, (2004)
3. Karagiannis, T., Papagiannaki, K., Faloutsos, M.: BLINC: Multilevel Traffic Classification in the Dark. In: SIGCOMM, (2005)
4. Dang, T. D., Perényi, M., Gefferth, A., Molnár, S.: On the Identification and Analysis of P2P Traffic Aggregation. In: Lecture Notes in Computer Science, Springer-Verlag, vol. 3976, pp. 606–617 (2006)
5. Choffnes, D., Bustamante, F.: Pitfalls for Testbed Evaluations of Internet Systems. In: ACM SIGCOMM Computer Communication Review, vol. 40, pp. 43–50 (2010)
6. Erman, D., Ilie, D., Popescu, A.: BitTorrent Session Characteristics and Models. In: Proceedings of the 3rd International Conference on Performance Modelling and Evaluation of Heterogeneous Networks, pp. 1–10, Ilkley, West Yorkshire, U.K. (2005)
7. Andrade, N., Neto, E.S., Brasileiro, F., Ripeanu, M.: Resource Demand and Supply in BitTorrent Content-Sharing Communities. In: Computer Networks, vol. 53, pp. 515–527 (2009)
8. Basher, N., Mahanti, A., Mahanti, A., Williamson, C., Arlitt, M.: A Comparative Analysis of Web and Peer-to-Peer Traffic. In: Proceedings of the 17th International Conference on World Wide Web, pp. 287–296, Beijing, China (2008)