Comparative Traffic Analysis Study of Popular Applications

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. The popularity of applications changes fast in current Internet and the result is that the characteristics of Internet traffic also goes over a rapid change. In this paper we present the main characteristics of three popular applications based on actual measurements taken from a commercial network. BitTorrent as one of the leading P2P file sharing applications, YouTube as the head of video sharing applications and Facebook as the prominent online social networking application are investigated. Comparative results at both application- and flow-levels are presented and discussed.

Keywords: traffic measurements, traffic analysis, BitTorrent, YouTube, Facebook

1 Introduction

The fact that the Internet is a fast evolving world is manifested in the incredible quick change of applications. Over the years we can observe a dramatic change in the popularity of applications used in the Internet. After the period when the *Web* was the leading application we could identify the dominance of *peer-to-peer file sharing* applications for a while. A number of P2P applications have been developed and a few of them, e.g. *BitTorrent* reached extremely high popularity. However, it seems that after the P2P file sharing era, in recent years, there are some other increasing applications, which are becoming very popular such as *online social networking*. In this so-called Web 2.0 world the user-generated content sites provide platforms for information, video and photo sharing as well as blogging.

As an example, *YouTube* is the leading video sharing application with continuously increasing popularity. It was launched in December 2005, and since July 2006 the site serves up to 100 million videos per day with a daily upload of more than 65000 videos and nearly 20 million unique visitors per month [1]. The success of YouTube is interesting to investigate. The site exerts no control over its users' freedom for publishing, so users not only share their videos, but also participate in a huge decentralized community by creating and consuming terabytes of video content.

Nowadays, online social networking sites are also popular. The primary purpose of these sites is to provide the means for users to maintain contacts, communicate and exchange information with each other. The most prominent example is *Facebook*, which is a web-based online social networking application. Facebook is quickly emerging and can be seen as a new Internet killer-application.

In this paper we investigate the main traffic characteristics of three popular applications of different types: BitTorrent as one of the leading P2P file sharing applications, YouTube as the head of video sharing applications and Facebook as the prominent online social networking application. We carried out a comprehensive analysis study based on measurements taken from a commercial network. We analyzed and compared the traffic characteristics of these successful applications at both application- and flow-levels.

The paper is organized as follows. Section 2 overviews the related work, and in Section 3 we discuss the details of measurements including the network architecture and analysis tools. Section 4 and Section 5 present our analysis results at application- and flow-levels, respectively. Finally, Section 6 concludes the paper with our main results.

2 Related Work

The popular Internet applications are in the focus of several recent studies from different aspects. Since in our paper we address BitTorrent, YouTube and Facebook applications, we overview some related work in this section.

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 [2]. 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 [3]. 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 [4].

YouTube patterns are intensively investigated in [5], and one of the main findings is that caching could improve the user experience, reduce bandwidth consumption and reduce the load on YouTube servers. Zink et al. analyzed YouTube traffic in a large university campus network [6]. They showed that there is no strong correlation between local and global popularity of YouTube videos. They also observed that many users watched the same video more than once. Cheng et al. presented a measurement study of YouTube videos and revealed that YouTube has a significantly different statistics compared to other video streaming applications especially in length distribution, access pattern and growth trend [7].

The "Facebook phenomenon" has been investigated in many recent papers from different points of view including both social [8] and technical aspects [9], [10]. Gjoka et al. carried out a measurement-based characterization of the popularity and usage of Facebook applications [9]. They found that the popularity of these applications is a rather skewed distribution. The authors of [10] also made a large-scale measurement study and they analyzed the usage characteristics of online social network based applications. They pointed out that only a small fraction of users account for the majority of activity within the context of Facebook applications.

3 Traffic Measurements

Measurements were taken from one of the commercial networks in Stockholm, Sweden. This company maintains network infrastructure for several service providers, which offer many different services such as Internet access, IP telephony or IPTV for both residential and business users. During the measurement period more than 1800 customers used the network for their own purposes.

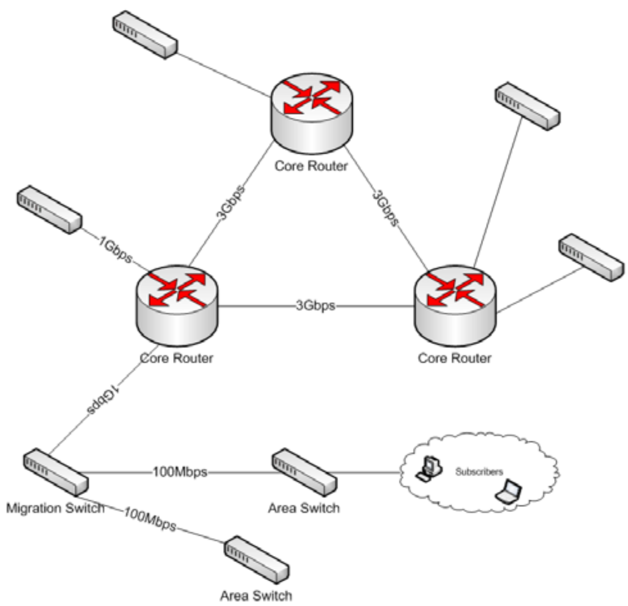


Fig. 1. The architecture of the network

The network infrastructure of Swedish backbone network and the related residential network are shown in Fig. 1.

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.

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. 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

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.

4 Application-level Analysis

Firstly, this section presents some important properties of the measured traffic including the daily profile and user ranking based on the number of incoming and outgoing packets. After that the main characteristics of the three chosen applications are investigated.

The daily profile of the traffic is given in Fig. 2 using the time interval (7) 12:00 – (8) 11:59. As it can be seen, the peak hours are between 5 PM and 8 PM. We can observe that the traffic volume generated in this time frame is approximately six times higher than in the morning hours.

Fig. 3 shows the measured incoming (downlink) packets as a function of the measured outgoing (uplink) packets. Every point in the figure corresponds to

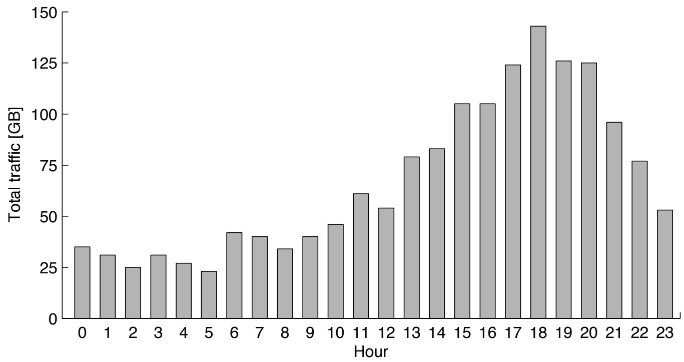


Fig. 2. Daily profile (FL-1)

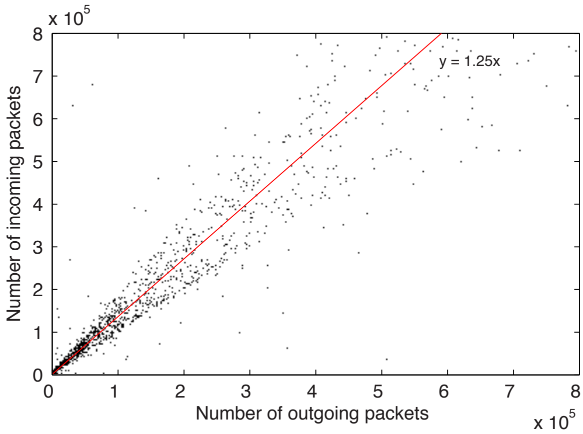


Fig. 3. User ranking based on the number of incoming and outgoing packets (FL-1)

one user. The linear fitting tells that an average user generates approximately 25% more downlink packets, though the variance is quite large.

Table 2 represents the incoming, outgoing and total traffic volumes generated by BitTorrent, YouTube and Facebook applications. The main difference among the applications is that while the amount of data uploaded by BitTorrent users is more than the downlink traffic volume (68% of the total BitTorrent traffic is uplink traffic and this is two times higher amount of data compared to the BitTorrent downlink traffic), YouTube and Facebook users primarily generate 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. In case of YouTube downloading takes 98% of the total traffic that can be explained by the specific characteristics of this application. Namely, video streaming produces a huge downlink traffic while uploading is necessary only for communicating with the servers and exchanging basic information. Although, users can upload videos that increases the uplink traffic, most of them consume only the contents uploaded by others. Our analysis of YouTube is based solely on the video contents, other user activities such as browsing and searching were excluded. In the whole measurement period (FL-1) users downloaded from 1972 different YouTube servers while in the selected busy hour (FL-2) this value is 222. In contrast to the previous results, Facebook generates only a small amount of network traffic.

Table 2. The traffic volume generated by the investigated applications (FL-2)

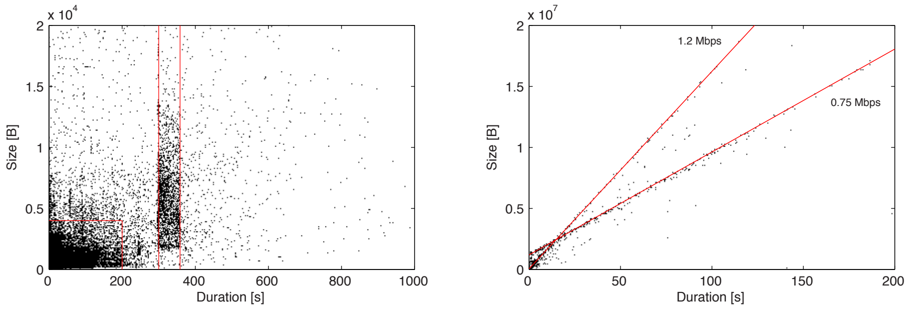
Application	**Incoming**	**Outgoing**	**Total**
BitTorrent	7.5 GB (32%)	15.8 GB (68%)	23.3 GB
YouTube	3.62 GB (98%)	63 MB (2%)	3.68 GB
Facebook	45 MB (79%)	12 MB (21%)	57 MB

Concerning user penetration 13%, 3% and 6% of the active customers used BitTorrent, YouTube and Facebook applications, respectively. However, these values are not easy to compare for different applications. For example, while BitTorrent clients often run all day long, an average user visits Facebook once a day.

5 Flow-level Analysis

In the following analysis both incoming and outgoing traffic have been considered and results are related to the total traffic. In this about one and a half day long period 1217 GB was downloaded (incoming traffic) and 1568 GB was uploaded (outgoing traffic). It clearly shows 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 illustrates the relationships between flow size and duration for BitTorrent and YouTube applications in an enlarged view. Every point in Fig. 4 and Fig. 6 represents exactly one flow. There are two different clusters in Fig. 4a: a cluster concentrated around the horizontal and vertical axis, and another one bounded by the vertical lines. Our analysis showed that the number of bytes carried by a BitTorrent flow is 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. 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. Fig. 4b shows that most of the YouTube flows are located along two lines having different gradients. In other words, the flow rate primarily varies around 1.2 Mbps and 0.75 Mbps, respectively.



(a) BitTorrent (b) YouTube

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

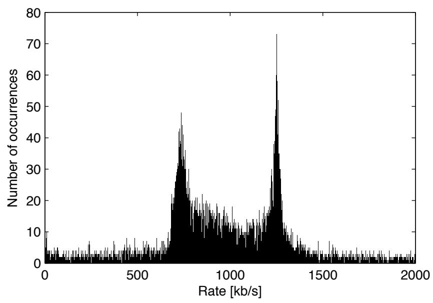
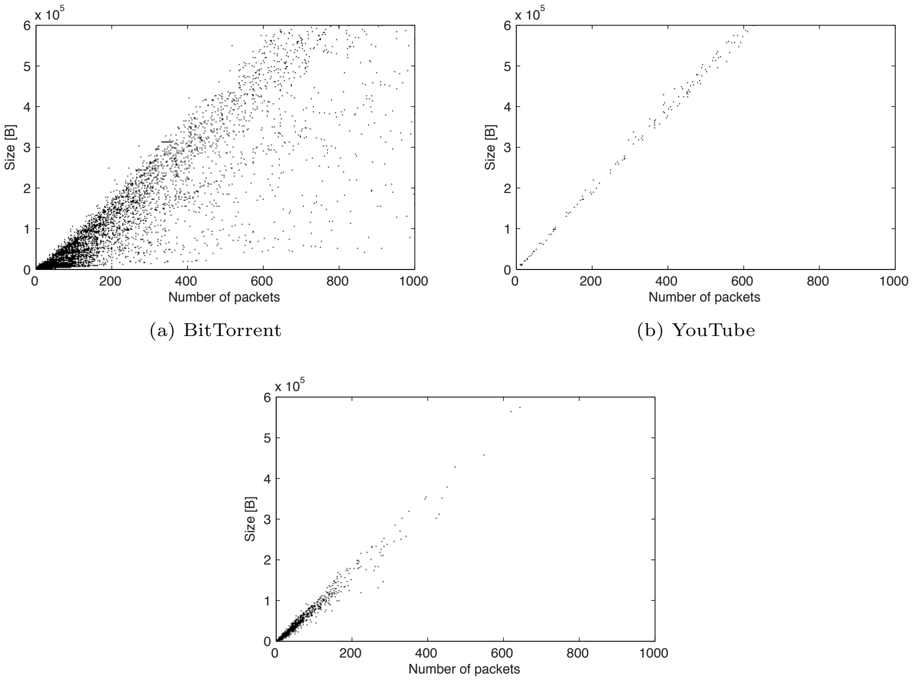


Fig. 5. Histogram of the flow rate for YouTube (FL-2)

This property can also be observed in the histogram of the flow rate depicted in Fig. 5. The first peak around 1.2 Mbps can be explained by the rate limitation of the YouTube servers and the second peak about 0.75 Mbps is due to YouTube control that balances the quality and the bandwidth.

Fig. 6 depicts the relationships between flow size and number of packets for the three popular applications. The size of the largest packet is near to the maximum size of the Ethernet frame close to 1500 bytes. In Fig. 6a and 6b it can be observed that BitTorrent flows are more scattered than YouTube and Facebook flows. It indicates that there is a negligible difference among the packet sizes in case of YouTube and Facebook, and the Ethernet frame size is better utilized compared to BitTorrent.

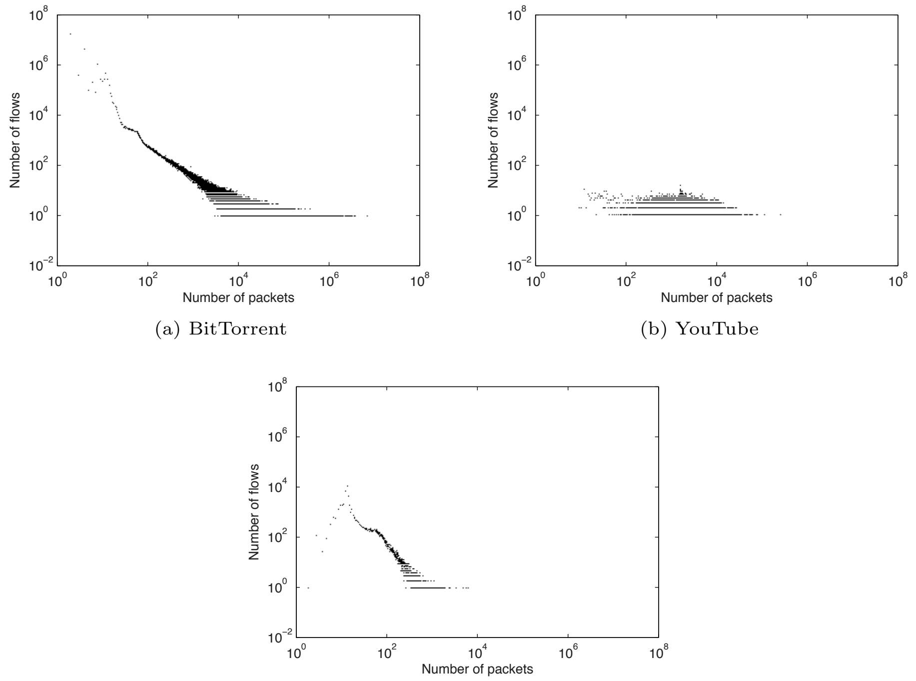


(a) BitTorrent (b) YouTube

(c) Facebook

Fig. 6. Relationships between flow size and number of packets (FL-2)

Fig. 7 shows the histograms of the number of packets for BitTorrent, YouTube and Facebook applications. We can observe that the number of BitTorrent and Facebook flows have a heavy-tailed decrease for the number of packets more than 100 as shown in Fig. 7a and 7c, respectively. In contrast to BitTorrent the histogram of YouTube flows has a unique characteristics (see Fig. 7b). There are only a few occurrences of a particular number of packets. A deeper analysis revealed the interesting property that in case of BitTorrent only 0.1% of flows contain unique number of packets. In contrast, most of the YouTube flows are unique since 57.4% of flows consist of different numbers of packets, but in case of Facebook flows this ratio is about 1.2%. The previous values were calculated for the whole measurement period (FL-1), but for the one hour long trace in the busy period (FL-2) we got the more surprising 0.4%, 93.6% and 8.9%, respectively. In this case the interesting property of YouTube is more pronounced: almost every YouTube flow contains different numbers of packets.



(a) BitTorrent (b) YouTube

(c) Facebook

Fig. 7. Histograms of the number of packets (FL-1)

6 Conclusion

In this paper we presented a comparative traffic characterization study of three popular applications based on actual measurements taken from a commercial network. BitTorrent as one of the leading P2P file sharing applications, YouTube as the head of video sharing applications and Facebook as the prominent online social networking application were investigated. We studied the main traffic characteristics of the measured data including the daily profile and the incoming and outgoing traffic ratios. We found that the number of incoming packets is 25% more than the number of outgoing packets, but the traffic volume shows a different picture: the outgoing traffic volume is about 30% higher than the incoming traffic volume. This observation can be 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.

The flow-level analysis revealed that the number of bytes carried by a BitTorrent flow is almost independent of the flow duration, and almost all of the BitTorrent 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. However, in case of YouTube we got a completely different result: flow sizes and durations are depending on each other. The typical rate of a YouTube flow varies around 1.2 Mbps and 0.75 Mbps. This is due to the rate limitation of the YouTube servers and the control that balances the quality and the bandwidth. Although, the number of BitTorrent and Facebook flows have a heavy-tailed decrease for the number of packets more than 100, almost all of the YouTube flows contain different numbers of packets. We clearly identified that YouTube has a unique characteristics in contrast to other applications such as BitTorrent and Facebook.

Acknowledgement. We thank Sollentuna Energi AB for the measurements, Ericsson Sweden and Ericsson Hungary for the cooperation and helping to access the data.

References

1. Reuters: YouTube serves up 100 million videos a day online. USA TODAY, `http://usatoday.com/tech/news/2006-07-16-youtube-views_x.htm` (2006)
2. Choffnes, D., Bustamante, F.: Pitfalls for Testbed Evaluations of Internet Systems. In: ACM SIGCOMM Computer Communication Review, vol. 40, pp. 43–50 (2010)
3. 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)
4. 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)
5. Gill, P., Arlitt, M., Li, Z., Mahanti, A.: YouTube Traffic Characterization: A View From the Edge. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, pp. 15–28, New York, NY, USA (2007)
6. Zink, M., Suh, K., Gu, Y., Kurose, Y.: Characteristics of YouTube Network Traffic at a Campus Network – Measurements, Models, and Implications. In: Computer Networks, vol. 53, pp. 501–514 (2008)
7. Cheng, X., Dale, C., Liu, J.: Understanding the Characteristics of Internet Short Video Sharing: YouTube as a Case Study. In: Proceedings of the 16th IEEE International Workshop on Quality of Service, Enschede, Netherlands (2008)
8. McClard, A., Anderson, K.: Focus on Facebook: Who Are We Anyway? In: Anthropology News, vol. 49, pp. 10–12 (2008)
9. Gjoka, M., Sirivianos, M., Markopoulou, A., Yang, X.: Poking Facebook: Characterization of OSN Applications. In: Proceedings of the 1st Workshop on Online Social Networks, pp. 31–36, Seattle, WA, USA (2008)
10. Nazir, A., Raza, S., Chuah, C.: Unveiling Facebook: A Measurement Study of Social Network Based Applications. In: Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement, pp. 43–56, Vouliagmeni, Greece (2008)