Unmasking the Growing UDP Traffic in a Campus Network

Changhyun Lee, DK Lee, and Sue Moon

Department of Computer Science, KAIST, South Korea

Abstract. Transmission control protocol (TCP) has been the dominating protocol for Internet traffic for the past decades. Most network research based on traffic analysis (e.g., router buffer sizing and traffic classification) has been conducted assuming the dominance of TCP over other protocols. However, a few recent traffic statistics are showing a sign of significant UDP traffic growth at various points of Internet links [26]. In this paper we show that the UDP traffic has grown significantly in recent years on our campus network; we have observed a 46-fold increase in volume (from 0.47% to 22.0% of total bytes) in the past four years. The trace collected in 2011 shows that the grown volume is not from a small number of UDP hosts nor port numbers. In addition, the recent UDP flows are not sent at constant bit rate (CBR) for most cases, and the aggregated traffic shows burstiness close to TCP traffic.

1 Introduction

Transmission control protocol (TCP) has been the main protocol of Internet traffic for the past decades; the widely accepted notion is that TCP accounts for more than 90% of the total traffic. User datagram protocol (UDP), on the other hand, has consumed only a small share of Internet traffic as it has been mainly used for limited purpose such as online gaming and multimedia streaming. Hence network engineering research has been based on the dominance of TCP traffic [5–7]. Traffic classification has also concentrated on identifying TCP applications, and only a few popular UDP applications such as PPLive and SopCast have been studied [8, 9]. In addition, network experiments with synthetic traffic have mostly focused on generating realistic TCP traffic while they often model UDP traffic as simple packet bunches sent at constant bit rate [24].

Recently, a few traffic statistics are showing the sign of UDP traffic growth at various points of Internet links [26]. The reported trend has not been studied thoroughly yet, and the cause and the impact of growing UDP traffic to the Internet are to be discovered. Although most traffic measurement studies have been about TCP, some previous research papers have looked at the characteristics of UDP traffic in terms of size, arrival, port usage of flows [17, 20, 22, 25]. However, the traffic traces used in those papers do not reflect the most recent trend as they are all collected before 2009 when only a small portion of UDP traffic around 5% or even less is reported.

In this work we report on the excessive growth in UDP traffic by continuous monitoring of the same network link for four years. We show the contribution of UDP to the overall traffic is no longer negligible according to the measurements from our campus

network; we have witnessed a 46-fold increase in volume (from 0.47% to 22.0% of total bytes) for the past four years. With the UDP trace collected in 2011, we characterize the UDP volume growth in terms of flow size, communication pattern, and sending rate. Here we refer to the total number of bytes as volume. Our results show that the growth in volume is mainly from the increase in the flow size rather than in the number of flows, and most UDP flows are not sent at constant bit rate (CBR). We have also found that the growth is not attributed to a small number of UDP servers nor port numbers. UDP is used by peer-to-peer file transfers today. Finally, we show the recent UDP traffic has comparable burstiness to that of TCP traffic.

The rest of this paper is organized as follows. Section 2 provides the data sets used in this work and evidences of recent growth in UDP traffic in terms of absolute volume, flow size, and packet size. In Section 3, we observe the sign of peer-to-peer applications on UDP by analyzing port usage and communication patterns between hosts. We then study the rate variation of UDP flows and burstiness of UDP traffic in Section 4, Last, Section 5 concludes with the implications and lessons from our findings.

2 Growth trend of UDP traffic

2.1 Data sets

We have collected the packet-level traces from 2008 to 2011 and captured the growth trend of UDP traffic on our campus network link. KAIST has a population of about 10,000, faculty, staff, and students all included and it owns 2 /16 prefixes and 80 /24 prefixes. A nearby college of about 1,100 got merged with KAIST in 2009 and KAIST acquired 1 of the 2 /16 prefixes and another 1 Gbps link to the outside. The campus network was reorganized in September 2009 that all traffic from the dormitories was routed via the new link and the rest of the configuration has remained almost the same. From 2008 to 2011, the overall population of KAIST grew from 7,000 to 10,000, mostly from the merger and the increase in the incoming student body size. Even with the increase in the overall population and network capacity, KAIST has not changed the traffic filtering policy: ICMP packets are dropped at the gateway but no traffic suspected to be peer-to-peer downloads.

We use GPS-synchronized servers with DAG 4.3GE cards [1] and collect header-only traces from the 1 Gbps link that connects classrooms, labs, and offices to a commercial ISP; we were not able to collect payload information due to the privacy concern in our campus.

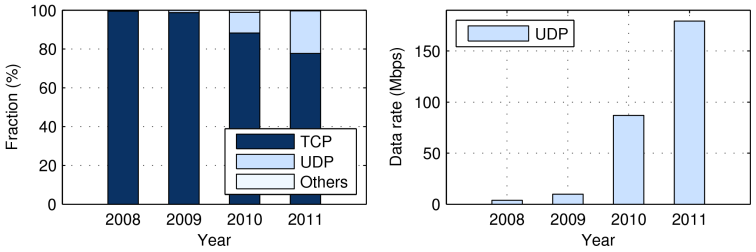
Trace name	Time of collection	Duration	Data rate
k-2008	2008/03/19 Wed 14:00	60min	937.2Mbps
k-2009	2009/04/27 Mon 14:00	60min	927.8Mbps
k-2010	2010/08/31 Tue 14:00	60min	868.5Mbps
k-2011	2011/01/07 Fri 14:00	60min	855.8Mbps

Table 1. Collected packet traces from 2008 to 2011

Table 1 shows the summary of collected traces used in this paper. The traces from 2008 to 2011 are all collected on weekdays and captured at the same time of the day to minimize errors from the diurnal effect in Internet usage. Traces *k-2008* and *k-2009* are before the merger and network reconfiguration and include traffic from the dormitories. The slight decrease in the overall data rates in 2010 and 2011 is attributed to the extra network capacity, but the link is still quite heavily utilized. In the spring of 2011 KAIST added another 1 Gbps link to the Internet.

In the rest of this paper we use the incoming traffic from the Internet core to KAIST to represent end-users' Internet usage.

2.2 Growth in UDP traffic volume



(a) Protocol breakdown (b) Data rate

Fig. 1. UDP traffic growth from 2008 to 2011

In Figure 1(a), we have found that UDP traffic has increased up to 22% over the past four years; the minimum among the traces is 0.47% in 2008. Surprisingly, the absolute volume of UDP traffic has grown significantly from 3.90 Mbps to 179.39 Mbps (46-fold growth in four years) in Figure 1(b). We see no sign of letting up in the UDP traffic growth. The growth trend is in compliance with a previous report [26], and we also find the similar trend from a trans-Pacific link of Japanese backbone networks [2] and CAIDA's two monitors in Chicago [4]. The UDP byte ratio of the Japanese traces from 2006 to 2011 has been growing, and the largest portion observed is around 30%. CAIDA's Chicago monitors report 15 18% in the average UDP data rate in August of 2011, which is about 5% higher than the recent two year's average on the same links. Although we show a single one-hour trace in 2011 in this paper for representation, the other traces collected in 2011 have similar shares of UDP around 20%, which is much larger than the share in 2008.

The number of UDP flows within an hour has also grown from 2.6 million in 2008 to 5.2 million in 2011, but not as much as in volume; we identify a UDP flow as a set of packets that have the same source and destination IP addresses and port numbers. The increase in the number of flows from 2008 to 2011 is only 2-fold. The more critical cause of the recent UDP traffic growth is the change in the size of each flow than in the total number of flows; the average flow size is 0.71 KB in 2008 and 16.32 KB in 2011. Previous work on TCP traffic trend has shown that the TCP flow size distribution has

remained similar from 2001 to 2009 [21], and our result on UDP here is in contrast to their finding. We give more details on the flow size evolution of recent UDP traffic in the next section.

2.3 Growth in UDP flow size

The common perception about UDP flows is that they are small and short, and it is supported by previous studies on the flow size of TCP and UDP [20, 26, 27]. We seek to verify whether it still holds for the recent UDP traffic. Figure 2 shows the cumulative volume by the flow size for the UDP traffic in trace *k-2011*. The figure also includes the distribution for UDP traffic in trace *k-2008* and TCP traffic in *k-2011* for comparison. We find that, in *k-2011*, flows larger than 100 KB take up 97.5% of the total volume. The same analysis on TCP traffic has shown 91.2%. Large flows dominate in UDP traffic as much as in TCP traffic or even more in some traces.

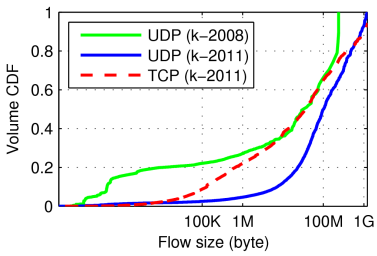


Fig. 2. Cumulative volume vs. flow size

The packet size of UDP traffic has also grown dramatically from 2008 to 2011. Figure 3 shows the cumulative distribution of packet size over four years. The portion of UDP packets larger than 1,400 bytes is 43.2% in *k-2011*, while only 0.34% in *k-2008*. We have found an interesting trend that the packet size distribution of UDP from 2008 to 2011 has become bimodal like that of TCP; in *k-2011*, packets either smaller than 100 bytes or larger than 1,400 bytes contribute 89.2% of the total packets for UDP and 91.5% for TCP.

Last we look at the duration of UDP flows. A flow's duration is calculated as the time between the first and the last packet within a flow. We find that 76.4% of flows have zero duration because they consist of only one packet. The lifetime of the flows with more than two packets spans up to one hour, an upper limit imposed by our data collection. In summary, there are a number of very short UDP flows of one packet, but a small number of large flows take up most of the volume.

3 UDP for P2P

In this section we take a close look at the UDP traffic from the most recent trace *k-2011*. We analyze port number usage and communication patterns between hosts, and

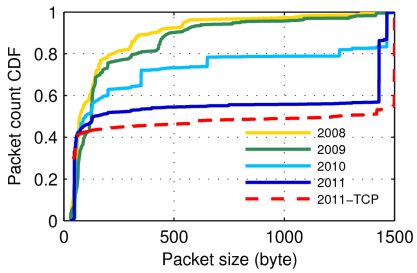


Fig. 3. Evolution in UDP packet size from 2008 to 2011 (top to bottom)

investigate the types of traffic contributing to the recent growth in UDP. From the result, we provide evidences that UDP is now used for peer-to-peer applications.

3.1 Port usage in UDP traffic

Port number usage is one of the key measures to understand the type of traffic and often used for identifying applications such as web surfing, online gaming, and peer-to-peer transfer with fixed port numbers [13, 15, 17, 22, 26]. Figure 4 plots the cumulative UDP traffic volume against the source and destination ports in *k-2011*. First, the source port numbers used by UDP flows are distributed all over the port allocation range. The largest volume on a single port number is 2.15% at the port $47{,}391$. The volume 2.15% is not so high compared to TCP traffic as it carries much volume on port 80 (HTTP), and traces from various network links report up to 62.9% of the total volume from HTTP [10, 15, 19]. We define a popular port as the port having more than 0.0015% out of the total volume; the threshold 0.0015% is set to the expected volume per port if traffic is distributed evenly over the port numbers. For the source port case, there are $2,496$ popular ports, and they account for 95.97% of the total traffic volume. Other than the well-known port 53 is used by queries to the DNS servers in KAIST, we are not able to map the popular port numbers to known applications only with the packet headers.

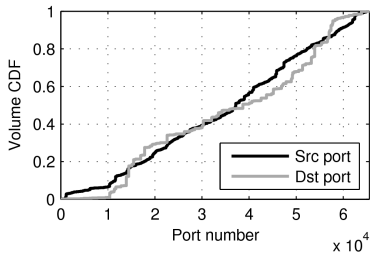
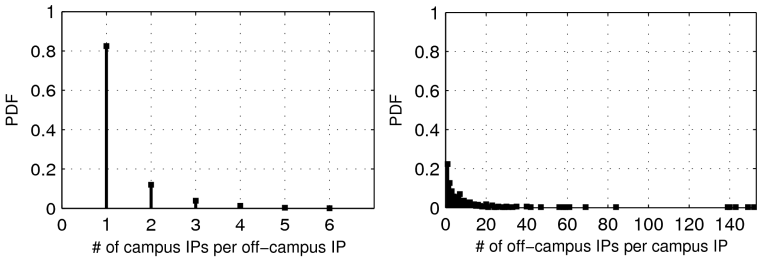


Fig. 4. Cumulative volume by the port number in *k-2011*

We apply the same analysis on destination ports. There are 594 popular ports and they are responsible for 99.23% of total traffic. The port $53{,}952$ carries the largest per-flow volume of 5.42%. Compared to the source port case, only a quarter of destination ports carry more traffic. Out of 594 ports 546 has 99% of volume coming from single nodes. Each of these nodes has a large number of flows up to thousands with the same destination port number but with different source port numbers. That is, a single destination port use used for multiple downloads of heavy volume on a single node. We investigate further the traffic distribution by the host in the next section.

3.2 Communication patterns between hosts



(a) Off-campus host popularity (b) Campus host popularity

Fig. 5. Communication pattern of UDP flows

Our next interest is the communication pattern between hosts of UDP flows. Analyzing the communication patterns, we try to determine if the recent UDP traffic comes from peer-to-peer type (many-to-many) transfers or server-to-client type (one-to-many) transfers. We take a similar methodology used in Karagiannis *et al.*'s work [14]. From the flow records, we first count the number of unique campus IP addresses per off-campus IP address. In the rest of this section we count only those flows larger more than 100 KB in the analysis. The threshold of 100 KB is arbitrary, but insensitive enough to exclude DNS and scanning traffic and to capture bulk transfers. As shown in Figure 5(a), most off-campus hosts (82.5% of the total) have only one corresponding host on campus, and the maximum number of corresponding hosts on campus is six. It means that no popular UDP source host exists outside for hosts on campus, and the growth in UDP traffic is not attributed to a single or several numbers of UDP off-campus servers. On the other hand, the same analysis on TCP traffic shows that the most popular server has sent traffic to 203 on-campus IPs within an hour in the same trace.

Figure 5(b) shows the number of unique off-campus IP addresses per campus IP address in the flow records, and the number goes up to more than hundred. Remember that all flows are larger than 100 KB here. That is, hosts on campus download UDP traffic from a large number of hosts outside. From the communications patterns by the host and the port in Section 3.1 we conclude that most UDP traffic is from peer-to-peer transfers than server-to-client transfers.

4 Burstiness of UDP Traffic

UDP traffic has increased to take up almost 20% of the total link capacity on our very congested link. If it is constant bit rate, not adaptive or responding to the network congestion, it would be equal to decreasing the available bandwidth and have an unfair share of bandwidth over TCP flows. When UDP is relatively a negligible portion of the overall traffic, this unfair advantage is not very important. Now it is an issue.

We use the standard deviation in flow throughput to first see if UDP flows are CBR or not. We count the number of bytes delivered in a time unit of one second and calculate its standard deviation per flow. We compute the same for TCP flows to compare with. Figure 6 shows the cumulative distributions of the standard deviation. As in the previous section all flows accounted for in this section are larger than 100 KB. In *k-2008* most UDP flows have zero standard deviation. However, as time progresses to 2011, UDP traffic shows an increasing tendency of variability in throughput. By *k-2011* about top 18% of both UDP and TCP flows have the standard deviation greater than 1.6 Mbps. The portion of UDP flows with almost zero variability drops to less than 30%.

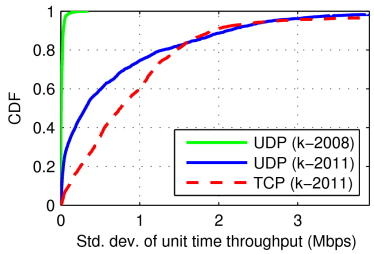
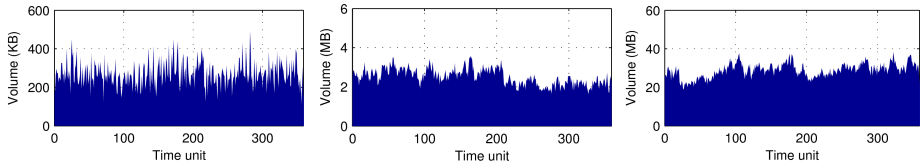


Fig. 6. Standard deviation of unit time throughput

Here we have looked at the throughput variability only in the time unit of a second and on a per-flow basis. One second is rather a long time for a router queue to buffer packets in today's Internet where most backbone links are 1 Gbps or higher: that is, too coarse a time scale. How variable or bursty is the aggregate UDP traffic in finer time scales? In Figure 7 we examine the burstiness of aggregate UDP traffic in time units of 0.01 s, 0.1 s and 1 s. At the time scale of 0.01 s the traffic looks more bursty than in the other two scales, but the other two look similar.

Burstiness in traffic has a great impact on router queue and end-host buffer size provisioning. Self-similarity in Internet traffic has long been reported and its causes have been studied [28] A common technique to analyze the scaling behavior in traffic is the wavelet analysis and its energy plot [29]. The energy plot in Figure 8 shows the variance of the wavelet coefficients that reflects the variance of traffic counting process X_j at a time scale T_j. If the traffic is self-similar, the plot should be a straight line, of which slope is the scaling exponent. If the traffic is Poisson, the plot should be a horizontal line.



(a) With 0.01s bin (b) With 0.1s bin (c) With 1s bin

Fig. 7. Burstiness in aggregate UDP traffic from *k-2011*

TCP traffic from *k-2011* in Figure 8 shows almost a straight line, signifying that it is close to self-similar. UDP traffic on the other hand has a slight tip near $j = 7$ or the time scale of 256 ms, where $j = 1$ is in 2 ms. The scaling exponent (or the Hurst parameter) for TCP is 0.865 and for UDP 0.831. Multi-scaling behaviors on high-speed links and similar dips in the time scale of hundreds of milliseconds have been reported [30]. We have no basis to imply that two dips have a common cause and leave it for future work.

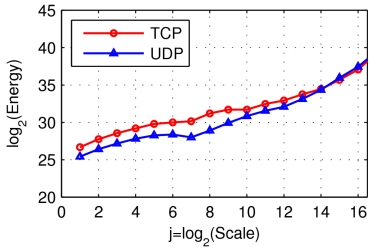


Fig. 8. Wavelet energy plot for TCP and UDP traffic from *k-2011*

Burstiness in UDP traffic does not instantly translate to use of congestion control by the applications. MPEG-coded video can by itself be bursty. However, the communication patterns of UDP hosts imply peer-to-peer transfers and the latest version of μtorrent, a popular client of BitTorrent, has announced the use of a proprietary congest control mechanism [3]. Our experiment of a μtp transfer on a controlled node shows that most data packets has a signature size of $1{,}466$ bytes, and we have identified 26.8% of total UDP traffic volume in *k-2011* to have the packet size. This is an upper bound as we may have false positives in our classification. The large volume of UDP flows with a proprietary congestion control contributes to the new kind of burstiness in today's Internet traffic.

5 Conclusions and discussion

In this work we have shown that UDP traffic has increased 46-fold over past four years on our campus network. Using packet header traces, we give a first characterization

report on the growth. From the trace collected in 2011, we have found that large flows have become dominant in UDP just as in TCP. They are mostly from P2P applications, and the aggregate UDP traffic exhibits burstiness similar to TCP.

Our findings provide several guidelines to classifying UDP traffic. First, port-number based classification can hardly work on recent UDP traffic. Port numbers seem to be randomly assigned to flows. This is an opposite result to the previous work on TCP traffic [15]. However, a destination port, once assigned, is used for multiple downloads from different hosts and ports, just as in TCP-based peer-to-peer applications. Thus the communication patterns can be a clue as in [13, 14]. In addition, we have found that certain UDP packet sizes, e.g., 1,466 bytes in *k-2011*, is observed more frequently than others. Packet sizes of UDP packets can be a good signature in identifying UDP applications. This is hardly the case for TCP since applications all work under TCP's policy.

Our observation on UDP traffic growth has implications to network simulation and experiments. In previous network experiments with synthetic traffic, UDP flows have been generated in a simple manner of constant bit rate and often ignored for their minor volume. While this has been valid for traditional UDP traffic, our measurements show that the packet sending behavior is much more bursty than simple CBR. Our measurement analysis underlines the rising need to account for "lower-than best effort" traffic in realistic network simulation.

Acknowledgements. This research was supported by the KCC (Korea Communications Commission), Korea, under the R&D program supervised by the KCA (Korea Communications Agency) (KCA-2011-08913-05002).

References

1. Endace, {http://www.endace.com}
2. Samplepoint-F Traces from MAWI Working Group Traffic archive (2006-2011), {http://mawi.wide.ad.jp/mawi}
3. What is μTorrent's μtp?, {http://www.utorrent.com/help/documentation/utp}
4. CAIDA's Passive Network Monitor Statistics, {http://www.caida.org/data/realtime/passive/}
5. Appenzeller, G., Keslassy, I., McKeown, N.: Sizing Router Buffers. In: Proc. ACM SIGCOMM (2004)
6. Beheshti, N., Ganjali, Y., Ghobadi, M., McKeown, N., Salmon, G.: Experimental Study of Router Buffer Sizing. In: Proc. ACM SIGCOMM IMC (2008)
7. Dhamdhere, A., Jiang, H., Dovrolis, C.: Buffer Sizing for Congested Internet Links. In: Proc. IEEE INFOCOM (2005)
8. Finamore, A., Mellia, M., Meo, M., Rossi, D.: KISS: Stochastic Packet Inspection Classifier for UDP Traffic. IEEE/ACM Trans. Netw., 18, 1505–1515 (2010)
9. Fu, T., Hu, Y., Shi, X., Chiu, D., Lui, J.: PBS: Periodic Behavioral Spectrum of P2P Applications. In: Proc. PAM (2009)
10. Henderson, T., Kotz, D., Abyzov, I.: The Changing Usage of a Mature Campus-wide Wireless Network. In: Proc. ACM Mobicom (2004)

11. Jiang, H., Dovrolis, C.: Source-level IP Packet Bursts: Causes and Effects. In: Proc. ACM SIGCOMM IMC (2003)
12. Jiang, H., Dovrolis, C.: Why is the Internet Traffic Bursty in Short Time Scales? In: Proc. ACM SIGMETRICS (2005)
13. Karagiannis, T., Broido, A., Faloutsos, M., claffy, K.: Transport Layer Identification of P2P Traffic. In: Proc. ACM SIGCOMM IMC (2004)
14. Karagiannis, T., Papagiannaki, K., Faloutsos, M.: BLINC: Multilevel Traffic Classification in thee Dark. In: Proc. ACM SIGCOMM (2005)
15. Kim, H., Claffy, K., Fomenkov, M., Barman, D., Faloutsos, M., Lee, K.: Internet Traffic Classification Demystified: Myths, Caveats, and the Best Practices. In: Proc. ACM CoNEXT (2008)
16. Lee, C., Lee, D., Yi, Y., Moon, S.: Operating a Network Link at 100%. In: Proc. PAM (2011)
17. Lee, D., Carpenter, B., Brownlee, N.: Observations of UDP to TCP Ratio and Port Numbers. In: Proc. IEEE ICIMP (2010)
18. Leland, W., Taqqu, M., Willinger, W., Wilson, D.: On the Self-similar Nature of Ethernet Traffic. IEEE/ACM Trans. Netw., 2, 1–15 (1994)
19. Maier, G., Feldmann, A., Paxson, V., Allman, M.: On Dominant Characteristics of Residential Broadband Internet Traffic. In: Proc. ACM SIGCOMM IMC (2009)
20. Olivier, P., Benameur, N.: Flow Level IP Traffic Characterization. In: Proc. ITC (2001)
21. Qian, F., Gerber, A., Mao, Z., Sen, S., Spatscheck, O., Willinger, W.: TCP Revisited: A Fresh Look at TCP in the Wild. In: Proc. ACM SIGCOMM IMC (2009)
22. Rodrigues, L., Guardieiro., P.: A Spatial and Temporal Analysis of Internet Aggregate Traffic at the Flow Level. In: Proc. IEEE GLOBECOM (2004)
23. Rossi, D., Testa, C., Valenti, S.: Yes, We LEDBAT: Playing with the New BitTorrent Congestion Control Algorithm. In: Proc. PAM (2010)
24. Sommers, J., Barford, P., Greenberg, A., Willinger, W.: An SLA Perspective on the Router Buffer Sizing Problem. ACM SIGMETRICS Perform. Eval. Rev., 35, 40–51 (2008)
25. Thopmson, K., Miller, G., Wilder, R.: Wide-area Internet Traffic Patterns and Characteristics. IEEE Network, 11, 10–23 (1997)
26. Zhang, M., Dusi, M., John, W., Chen, C.: Analysis of UDP Traffic Usage on Internet Backbone Links. In: Proc. IEEE/IPSJ SAINT (2009)
27. Kim, M., Won, Y., Hong, J.: Characteristic Analysis of Internet Traffic from the Perspective of Flows. Elsevier Computer Communications, 29, 1639–1652 (2005)
28. Park, K., Willinger, W.: Self-Similar Network Traffic and Performance Evaluation. John Wiley & Sons, Inc., New York (2002)
29. Abry, P., Veitch, D.: Wavelet Analysis of Long-Range-Dependent Traffic. IEEE Trans. on Information Theory, 44, 2–15 (1998)
30. Zhang, Z., Ribeiro, V., Moon, S., Diot, C.: Small-Time Scaling Behaviors of Internet Backbone Traffic: An Empirical Study. In: Proc. IEEE INFOCOM (2003)