Analysis of UDP Traffic Usage on Internet Backbone Links

Min Zhang*, Maurizio Dusi[†], Wolfgang John[‡] and Changjia Chen*
*Beijing Jiaotong University, China. Email: mia.minzhang@gmail.com
[†]Università degli Studi di Brescia, Italy. Email: maurizio.dusi@ing.unibs.it
[‡]Chalmers University of Technology, Sweden. Email: wolfgang.john@chalmers.se

I. Introduction

It is still an accepted assumption that Internet traffic is dominated by TCP [1], [2]. However, the rise of new streaming applications [3] such as IPTV (PPStream, PPLive) and new P2P protocols (e.g. uTP [4]) that try to avoid traffic shaping techniques (such as RST packet injection) will increase the use of UDP as a transport protocol. Since UDP lacks any functionality to adapt to network traffic congestion, a substantial increase in UDP usage raises serious concerns about fairness and stability in the Internet.

The goal of this paper is to shed light on the assumption that TCP is still the dominant transport protocol on the Internet, as reported by e.g. Fomenkov et al. in 2004 [1] and John and Tafvelin in 2007 [2]. We evaluate the amount of UDP and TCP traffic, in terms of flows, packets and bytes, on traces collected in the period 2002-2009 on several backbone links located in the US and Sweden. According to our best available data, the use of UDP as a transport protocol has gained popularity recently, especially in terms of number of flows. Our first analysis suggests that most UDP flows use random high ports and carry few packets with little content (payload), consistent with its use as a signaling protocol for increasingly popular P2P applications [5]. Many such applications build overlay networks to exchange information about how to share specific (and typically large) files; UDP allows efficient establishment and maintenance of such an overlay network, while use of random ports evades detection by port-based traffic engineering or filtering techniques.

II. Datasets

For this study we analyzed traffic traces from backbone links in the United States and in Sweden. The data from Sweden was collected on an OC192 link inside the GigaSUNET network during 2006, and on an OC192 connection link of the current OptoSUNET network. Traffic data from GigaSUNET includes two traces of 20 minutes collected in April and November 2006, summing up to 9M flows carrying 422M IP packets and 294GB of data. Two samples of 20-minute each were collected from OptoSUNET in January and February 2009, and include 41M flows, 1100M packets and 657GB of data [6].

The data from the US was collected on an OC48 peering link for a large ISP and on an OC192 backbone link. Two 60-minute traces were collected on the OC48 link in August 2002 and January 2003. The OC48 traces include 105M flows, 1834M packets and 1105GB of data. Traces from the OC192 link are also 60-minute long samples, collected in 2008 and 2009, and consist of 379M flows, carrying 8434M packets and 4446GB of data in total. Further details about the datasets are available at the CAIDA webpage[1].

III. Analysis of UDP traffic

We used CoralReef[2] to extract TCP and UDP flows from our traces. Each flow record, defined by the five-tuple (source and destination IP, port numbers and protocol), includes the counts of packets and bytes exchanged.

In Table I we report the ratio between UDP and TCP traffic, in terms of packets, bytes and flows. The use of UDP as a transport protocol has rapidly increased from 2002 to 2009, although TCP sessions are still responsible for most packets and bytes. However, in terms of flows, UDP dominates: on OptoSUNET (2009) we observe 3x as many UDP flows as TCP flows. Note that the OptoSUNET data include a substantial portion of traffic on UDP port 53, due to the presence of a RIPE DNS server located inside SUNET, serving over 400 zones. Traffic coming from and to port 53 of this server cannot be considered native SUNET traffic and we filtered it out for this study.

Trace	Sample	UDP/TCP Ratio		
		packets	bytes	flows
CAIDA-OC48	08-2002	0.11	0.03	0.11
	01-2003	0.12	0.05	0.27
GigaSUNET	04-2006	0.06	0.02	1.06
	11-2006	0.08	0.03	1.45
CAIDA-OC192	06-2008	0.14	0.05	1.43
	02-2009	0.19	0.07	2.34
OptoSUNET	01-2009	0.21	0.11	3.09
	02-2009	0.20	0.11	2.63

Table I
Values of UDP/TCP ratio

A per-port analysis helped us infer the nature of the UDP flows. Figure 1 reports the CDFs of the port numbers used by UDP flows (x-axis in log-scale). For the 2002-2003

This study was performed while the three first-named authors visited CAIDA at UCSD (http://www.caida.org) under supervision of kc claffy.

[1] http://www.caida.org/data/passive/

[2] http://www.caida.org/tools/measurement/coralreef/

traces, around 40% of UDP flows run on ports below 1024, including DNS (port 53), NTP (port 123) and NetBios (port 137). Since 2003, usage of ephemeral ports (>1024) has become common. After 2003, virtually all UDP ports are used, and nowadays around 95% of the UDP flows run on ports >1024. According to a port-based classifier, besides DNS, NTP and NetBios the top-used ports in terms of UDP flows are the ones normally used by P2P applications, such as 4672 and 4665 (eDonkey), 6881 (BitTorrent), 6346 (Gnutella) and 6257 (WinMX).

We attribute the flows running on those ephemeral ports to P2P overlay signaling traffic rather than to bulk data transfers. Our analysis reveals that flows on the top-ten ranked ports generally carry fewer than 7 packets and about 10KB on average: larger UDP flows appear mainly in the older traces (2002-2003), suggesting a drift in usage of UDP toward small (signaling) flows (see Figure 2).

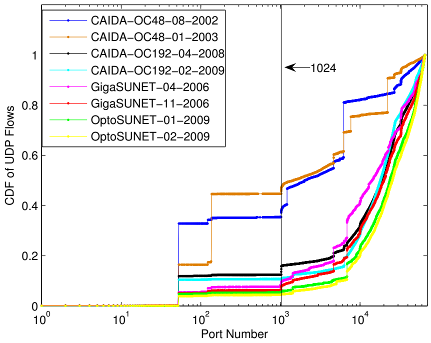

Figure 1. CDFs of UDP flows based on port number.

IV. Summary and Conclusion

We compared UDP and TCP traffic in several traffic traces collected from different networks and geographical locations, at different times. We find that TCP still dominates in terms of packets and bytes, but UDP is now often responsible for the largest fraction of flows on a given link. A port-based analysis suggests that the recent increase in UDP flows on the traces analyzed stems mainly from P2P applications using UDP for their overlay signaling traffic.

This trend may again change with the advent of IPTV and UDP based P2P applications, which not only signal, but also transport large data segments via UDP [3], [4]. We will continue to monitor available data to track trends in UDP usage, and specifically seek data from China where UDP-based IPTV traffic is already common. Finally, we note that precise traffic classification requires methods beyond simple port classification. Most current traffic classification techniques focus on TCP [7], [8], with only preliminary examination of techniques for UDP traffic [9] (other than deep packet inspection). Given the growing evidence for the use of UDP for increasingly popular applications, including for bulk data transfer in China, we conclude that traffic analysis methods must evolve to classify also UDP traffic.

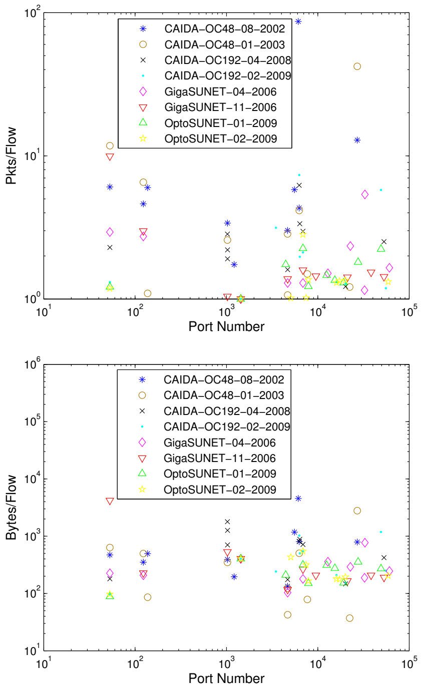

Figure 2. Average packets (top) and bytes (bottom) per flow, log-scale.

References

[1] M. Fomenkov, K. Keys, D. Moore, and K. Claffy, “Longitudinal study of internet traffic in 1998-2003,” in *WISICT*, 2004.

[2] W. John and S. Tafvelin, “Analysis of internet backbone traffic and header anomalies observed,” in *ACM IMC*, 2007.

[3] P. Pan, Y. Cui, and B. Liu, “A measurement study on video acceleration service,” in *IEEE CCNC*, 2009.

[4] Wikipedia.org, “Micro transport protocol,” Online: ”http://en.wikipedia.org/wiki/Micro_Transport_Protocol”, accessed April 29, 2009.

[5] W. John, S. Tafvelin, and T. Olovsson, “Trends and differences in connection-behavior within classes of internet backbone traffic,” in *PAM*, 2008.

[6] W. John and S. Tafvelin, “SUNET OC192 traces,” Online: ”http://imdc.datcat.org/collection/1-04HN-W=SUNET+OC+192+Traces”, accessed April 28, 2009.

[7] T. Nguyen and G. Armitage, “A survey of techniques for internet traffic classification using machine learning,” *IEEE Communications Surveys & Tutorials*, vol. 10, no. 4, 2008.

[8] M. Zhang, W. John, K. Claffy, and N. Brownlee, “State of the art in traffic classification: A research review,” *PAM Student Workshop*, 2009.

[9] T. Z. Fu, Y. Hu, D. M. Chiu, and J. C. Lui, “Pbs: Periodic behavioral spectrum of p2p applications,” *PAM*, 2009.