Observations of UDP to TCP Ratio and Port Numbers

DongJin Lee, Brian E. Carpenter, Nevil Brownlee
Department of Computer Science
The University of Auckland, New Zealand
dongjin, brian@cs.auckland.ac.nz, nevil@auckland.ac.nz

***Abstract*—Widely used protocols (UDP and TCP) are observed for variations of the UDP to TCP ratio and of port number distribution, both over time and between different networks. The purpose of the study was to understand the impact of application trends, especially the growth in media streaming, on traffic characteristics. The results showed substantial variability but little sign of a systematic trend over time, and only wide spreads of port number usage.**

***Index Terms*—network traffic; observation; ratio; port number**

I. Introduction

Along with annual bandwidth growth rates reported to be 50% to 60% per year both in the U.S. and worldwide [7], Internet traffic types, characteristics and their distributions are always changing. For example, a recent 2009 Internet Observatory report [18] finds that majority of traffic has migrated to a small number of very large hosting providers, such as those supporting cloud computing. Also, it has been widely predicted that within a few years, a large majority of network traffic will be audio and video streaming. Cisco's Virtual Networking Index [4] has been actively involved in traffic forecasting, e.g., *Hyperconnectivity and the Approaching Zettabyte Era* [5]. Those reports assert that by year 2010 video will exceed p2p in volume, and be the main source of future IP traffic growth. They also state that video traffic can change the economic equation for service providers, given that video traffic is many times less valuable per bit than other content such as SMS service. Additionally, increases in monitor screen size and its resolution give rise to larger document sizes (such as more pixels in images and videos), thus generating more traffic than before.

A common expectation in the technical community has been that streaming traffic would naturally be transmitted over UDP, probably using RTP, or perhaps in future over DCCP. Another view is that UDP and TCP might replace IP as the lowest common denominator [23] to achieve transparency through NATs and firewalls. Then, if non-TCP congestion control, signaling or other features are needed, a protocol must be layered on top of UDP instead of developing a better transport layer. This, if accompanied by a vast increase in streaming, would change the historic pattern whereby most traffic benefits from TCP's congestion management. Therefore, the evolution of the observed UDP to TCP ratio in actual Internet traffic is a subject of interest. Indeed, if the predicted increase in streaming traffic were to remove most flows from any form of congestion control, the consequences would be serious. The UDP to TCP ratio has been briefly observed by CAIDA [1], where UDP flows are often responsible for the largest fraction of traffic. Their summary indeed suggests that the current ratio can change with increasing demand for IPTV and UDP-based real-time applications. We note that audio/video 'streaming' is not really a well-defined term, and it covers a variety of technologies. In some cases, for example some video-on-demand solutions, packets are transmitted over TCP or even over HTTP. In others, for example some voice-over-IP solutions, streams are transmitted over UDP. Some streaming applications choose dynamically whether to use UDP, TCP or HTTP.

Our expectation was that the growth in streaming traffic would be reflected in a steady growth in the UDP to TCP ratio, or in a systematic change in the relative usage of various port numbers, or both. We conducted a preliminary survey on the basis of readily available data from a variety of measurements, in both commercial and academic networks, between 1998 and 2008. It showed that the UDP to TCP ratio, measured by number of packets, varied between 5% and 20%, but with no consistent pattern over the ten years. For Internet2, it was 0.05 in 2002, 0.22 in 2006, and 0.15 in 2008. Similar inconsistencies showed up in partial data from observations in Norway, Sweden [15], Japan, Germany, the UK, and elsewhere. These inconsistencies were surprising, and did not suggest a steady growth in UDP streaming. To better understand these issues, we observe how TCP and UDP traffic have varied over the years, either by number of flows, or by their volume/duration.

We consider this study to be valuable to the service providers and network administrators managing their traffic. This includes outlining statistical datasets and deriving strategies, such as classifying application types, prioritizing specific flow types, and provisioning based on usage scenarios. Also, a definite trend in the fraction of non-flow-controlled UDP traffic might affect router design as far as congestion and queue management is concerned. In this paper, we particularly observe two behaviors, 1) variation of UDP to TCP ratio over time, and 2) port number distribution. As far as is possible from the data, we also observe application trends. We use the term "flow ratio" and "volume ratio" to represent the ratio of $\frac{UDP}{TCP}$ for their flow counts and data volumes respectively.

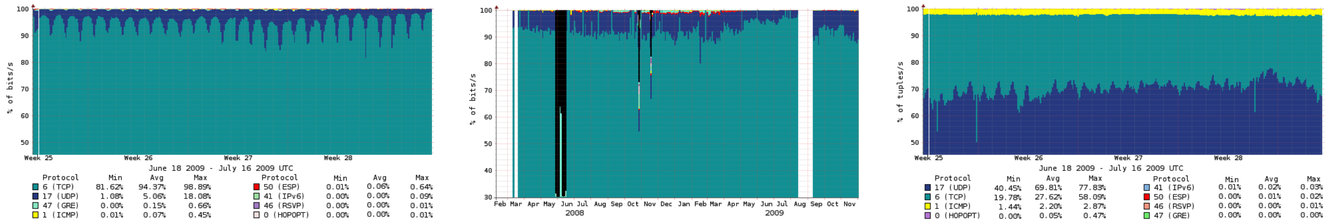

Fig. 1. CAIDA (2008–2009), Left: DirA – 4 weeks (bits), Center: Dir DirA – 20 months (bits), Right: DirB – 4 weeks (flows)

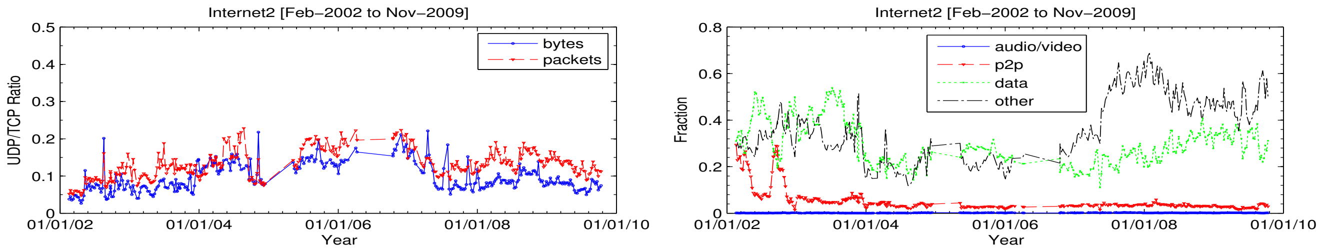

Fig. 2. Internet2 (2002-2009), Left: UDP to TCP ratio, Right: "audio/video", "p2p", "data" and "other" traffic volume

II. Longitudinal Data

Long term protocol usage is observed from two locations: CAIDA [2] and the Internet2 [6] monitor[1]. CAIDA traffic data is from the OC192 backbone link of a Tier1 ISP between Chicago and Seattle (direction A and B), reflecting various end-user aggregates. The Internet2 traffic reflects usage patterns by the US research and education community. Both datasets have HTTP and DNS traffic as the most widely used protocols for TCP and UDP respectively, but no particular specific application protocol was used predominantly.

Figure 1 shows plots for the CAIDA data. Although protocols such as ICMP, ESP and GRE are observed as well, TCP and UDP are in general most widely observed. We did not see a noticeable amount of SCTP or DCCP traffic. We observe that both DirA and DirB traffic contained about 95% TCP and 4% UDP bytes, measured daily and monthly (left and right). The volume ratio varied around an average of 0.05; the diurnal variation shows that during the peak time TCP volume (mainly HTTP) contributed as high as 98%, and during the offpeak time UDP volume can increase to 18%. Flow proportions (B, right plot) varied greatly as UDP flows are a lot more observed than TCP flows, e.g., on average 70% and as high as 77% of all flows are UDP. ICMP flows are stable, about 2%.

The dataset from Internet2 (Figure 2) covers a longer period of measurement, from February 2002 to November 2009. On left, we observe that the volume ratio has increased from early 2002 to mid 2004, then decreased from late 2006 to mid 2007, and again slight variations are observed from mid 2007 on. The UDP decrease observed in 2006 to 2007 may be due to the University of Oregon switching off a continuous video streaming service [14]. Generally the volume ratio varied between 5% and 20%, showing a higher variation than that of the CAIDA data. Comparing between 2002 and 2009, we find that the ratio of both bytes and packets has increased slightly by about 5%.

In this, there seems to be little evidence of change in protocol ratio, as most are diurnal variations with no particular increasing or decreasing patterns. On right, both audio/video and p2p traffic are little utilized over the period, whereas data (consisting mainly of HTTP traffic) and other (using ephemeral port numbers) traffic have increased. For example, audio/video traffic contributes to about 0.3% and p2p traffic decreased from about 20% to only about 2%. This could indicate that audio/video streaming and file sharing have genuinely decreased as compared to typical HTTP traffic, or that there are emerging applications using arbitrary port numbers or 'hiding' such traffic inside HTTP (e.g., [16]). Indeed, since about beginning of 2007, both the data and other traffic have increased substantially, from about 20% to more than 50%.

III. Port number

We next report observations from various different network locations measured in different years. Particularly, we observe port number distributions by using network traces[2] covering various network types. Table I shows a summary of measured traces. In total, 21 traces are so far measured by our traffic meter. A flow is identified by a series of packets with the same 5-tuple fields (source/destination IP address, source/destination port number, and protocol) and terminated by the fixed-timeout of 30 seconds. Since a flow is unidirectional, flow's source port number is used for observations.

[1] Note that the datasets contained some irregular anomalies throughout the period which have been removed from the plots. For example, short but very high peak usage of unidentified protocol, missing-data and inconsistent data values were observed and discussed with the corresponding authors at CAIDA and Internet2. They are presumed to be due to occasional instrumentation errors or, in some cases, to overwhelming bursts of malicious traffic. If included in the analysis, they would dominate the traffic averages and invalidate overall protocol trends. The original data including these anomalous peaks are available at the cited web sites.

[2] CAIDA [2], NLANR PMA [8] and WAND [10]

TABLE I
SUMMARY OF NETWORK TRACES

Trace Name	Network Type	Date, [Starting time], Duration (hours)	Average Rate (Mb/s)	Volume: Bytes (GB)	Volume: TCP (%)	Volume: UDP (%)	Volume: ICMP (%)	Volume: Other (%)	Volume: UDP/TCP Ratio	Number of Flows: Flows (M)	Number of Flows: TCP (%)	Number of Flows: UDP (%)	Number of Flows: ICMP (%)	Number of Flows: UDP/TCP Ratio
AUCK-99	UNIV	1999-Nov-29, [13:42], 24.00	1.39	14.96	94.26	5.51	0.19	0.04	0.06	2.63	82.52	15.32	2.17	0.19
AUCK-03	UNIV	2003-Dec-04, [00:00], 24.00	6.32	68.23	93.25	6.14	0.24	0.34	0.07	19.49	75.53	21.85	2.63	0.29
AUCK-07	UNIV	2007-Nov-01, [16:00], 24.00	60.41	652.41	94.70	4.72	0.43	0.15	0.05	73.62	44.44	52.73	2.82	1.19
AUCK-09	UNIV	2009-Aug-03, [09:00], 11.00	375.93	1860.85	93.77	6.12	0.02	0.08	0.07	93.84	59.65	39.45	0.90	0.66
BELL-I-02	ENT	2002-May-20, [00:00], 96.00	1.78	76.79	90.70	8.58	0.05	0.66	0.09	6.42	94.39	3.68	1.98	0.04
CAIDA-DirA-02	BB	2002-Aug-14, [09:00], 3.00	363.14	490.24	94.91	3.83	0.09	1.17	0.04	45.95	84.86	12.73	2.4	0.15
CAIDA-DirB-03	BB	2003-Apr-24, [00:00], 1.00	117.93	53.07	94.86	4.66	0.10	0.38	0.05	11.49	78.59	19.28	2.13	0.24
CAIDA-DirA-09	BB	2009-Mar-31, [05:59], 1.03	1250.83	579.76	96.69	2.74	0.48	0.09	0.03	46.96	43.16	54.46	2.38	1.26
CAIDA-DirB-09	BB	2009-Mar-31, [05:59], 1.03	3687.70	1709.25	91.17	8.11	0.06	0.66	0.09	61.03	32.50	65.06	2.44	2.00
ISP-A-99	COMML	1999-Nov-02, [14:04], 28.28	0.36	4.60	98.16	1.75	0.08	0.01	0.02	0.78	61.63	37.03	1.34	0.60
ISP-A-00	COMML	2000-Jan-04, [09:47], 32.80	0.37	5.44	94.37	5.44	0.08	0.12	0.06	0.94	57.86	40.68	1.46	0.70
ISP-B-05	COMML	2005-Jun-09, [07:00], 24.00	275.16	2971.74	92.26	6.93	0.22	0.59	0.08	513.76	62.88	33.79	3.32	0.54
ISP-B-07	COMML	2007-Feb-08, [00:00], 24.00	341.66	3689.90	94.43	5.05	0.12	0.40	0.05	500.56	49.61	46.35	4.05	0.93
LEIP-II-03	UNIV	2003-Mar-21, [21:00], 24.00	25.30	273.26	88.75	9.40	0.15	1.70	0.11	54.99	60.15	35.58	4.28	0.59
NZIX-II-00	IX	2000-Jul-06, [00:00], 96.00	3.50	151.38	87.35	9.23	3.39	0.03	0.11	55.28	47.18	29.88	22.94	0.63
SITE-I-03	ENT	2003-Aug-20, [04:20], 24.00	24.86	268.44	98.50	0.61	0.81	0.08	0.01	30.72	36.41	5.46	58.13	0.15
SITE-II-06	ENT	2006-May-11, [15:30], 33.90	76.52	1167.32	98.96	0.76	0.01	0.26	0.01	21.76	79.37	19.32	1.62	0.24
SITE-III-04	COMML	2004-Jan-21, [06:00], 24.30	110.15	1204.52	94.26	5.24	0.21	0.25	0.06	156.69	67.80	24.11	8.10	0.36
WITS-04	UNIV	2004-Mar-01, [00:00], 24.00	3.45	37.29	93.29	5.45	0.42	0.83	0.06	15.68	41.76	54.77	3.50	1.31
WITS-05	UNIV	2005-May-12, [00:00], 24.00	5.41	58.40	97.22	2.19	0.14	0.45	0.02	18.33	56.76	42.12	1.12	0.74
WITS-06	UNIV	2006-Oct-30, [00:00], 24.00	7.34	79.25	95.83	3.42	0.29	0.45	0.04	27.75	33.43	65.03	1.54	1.95

TABLE II
TOP-10 PORT USAGE

AUCK-09 - TCP				BELL-I-02 - TCP				CAIDA-DirB-09 - TCP				ISP-B-05 - TCP				LEIP-II-03 - TCP				NZIX-II-00 - TCP			
Flows		Volume		Flows		Volume		Flows		Volume		Flows		Volume		Flows		Volume		Flows		Volume	
Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%
80	34.89	80	70.41	80	28.35	119	32.28	80	24.41	80	65.58	80	6.90	80	16.17	4662	28.79	80	23.70	80	24.21	80	44.96
443	5.32	3131	5.99	2000	2.38	80	28.12	25	2.40	443	1.18	4662	3.46	4662	4.98	80	9.79	4662	9.00	443	2.09	20	2.96
3128	3.14	443	4.13	443	2.04	6677	2.59	9050	2.04	554	0.98	6881	2.30	6881	3.22	4661	0.81	6699	4.91	25	1.57	443	2.19
3131	1.38	3128	3.86	25	1.57	564	2.45	443	1.19	9050	0.84	6346	1.43	6346	2.93	443	0.46	1214	4.76	110	1.54	110	1.47
25	1.03	554	2.02	5190	1.34	10986	1.41	2710	0.45	81	0.39	25	1.18	8000	1.63	1214	0.41	2634	0.94	53	0.61	6699	1.30
1863	0.45	1935	1.08	21	1.31	22	1.29	445	0.34	1935	0.36	445	0.84	6699	1.15	6346	0.39	1755	0.90	3128	0.42	119	0.88
6000	0.37	993	0.31	22	0.99	554	1.20	6667	0.32	35627	0.19	1863	0.76	119	0.88	21	0.31	554	0.88	113	0.39	8080	0.87
2703	0.20	873	0.30	711	0.89	443	1.20	22	0.22	51413	0.13	16881	0.57	110	0.77	5190	0.30	20	0.58	2048	0.26	53	0.87
9050	0.20	22	0.17	1863	0.32	1755	1.02	11762	0.19	5001	0.11	110	0.56	6348	0.74	1841	0.26	22	0.56	20	0.23	4044	0.81
993	0.13	8002	0.11	5050	0.16	55418	0.98	21	0.17	52815	0.11	135	0.38	16881	0.56	25	0.26	2959	0.45	37	0.23	2048	0.75
Top10	47.11	Top10	88.38	Top10	39.35	Top10	72.55	Top10	31.72	Top10	69.87	Top10	18.37	Top10	33.04	Top10	41.77	Top10	46.69	Top10	31.54	Top10	57.07
Top20	47.77	Top20	89.19	Top20	40.23	Top20	79.05	Top20	32.76	Top20	70.78	Top20	20.36	Top20	36.13	Top20	43.32	Top20	50.20	Top20	32.63	Top20	60.01

AUCK-09 - UDP				BELL-I-02 - UDP				CAIDA-DirB-09 - UDP				ISP-B-05 - UDP				LEIP-II-03 - UDP				NZIX-II-00 - UDP			
Flows		Volume		Flows		Volume		Flows		Volume		Flows		Volume		Flows		Volume		Flows		Volume	
Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%	Port#	%
53	43.76	33001	24.69	137	21.41	7331	72.10	53	6.88	57722	2.56	4672	21.29	6346	8.59	4672	13.63	27015	17.59	53	32.41	27500	15.86
1513	0.92	33670	19.91	53	3.87	33264	2.79	6881	0.61	53	1.88	6881	8.14	6348	3.66	6257	4.56	27005	8.59	123	18.88	53	14.71
123	0.63	38168	7.91	123	3.33	161	2.57	6257	0.30	60096	1.32	53	6.79	7000	2.51	53	3.20	1701	3.71	1486	1.47	27005	9.46
14398	0.17	59002	5.34	32532	2.37	24716	2.22	6346	0.20	3074	1.25	6346	3.95	4672	2.48	1214	2.38	6257	2.39	4978	1.04	27015	5.59
17822	0.16	16402	4.58	500	1.35	53	1.59	45682	0.17	15000	1.22	6257	1.46	53	2.37	1841	2.15	27010	2.21	1553	1.03	27910	4.71
10306	0.15	53	3.55	24503	1.31	24504	1.17	60001	0.16	49262	0.98	123	0.98	16881	2.19	2857	1.28	53	1.52	4888	0.62	6112	4.18
36589	0.10	59004	1.96	27732	1.18	22888	1.06	32768	0.09	5004	0.56	1083	0.71	27005	1.87	3407	1.12	14758	1.18	137	0.57	123	1.85
51504	0.10	5442	1.89	6899	1.18	6899	1.01	50000	0.08	18350	0.47	6190	0.70	27016	1.50	3847	1.10	7714	0.98	1646	0.54	26005	1.44
2535	0.08	65321	1.58	55	1.14	7170	0.85	20129	0.08	4500	0.46	32770	0.68	6881	1.27	4964	1.09	3281	0.91	1024	0.54	28001	1.31
41048	0.08	1044	1.00	28753	1.02	137	0.81	60000	0.07	1044	0.46	1087	0.52	6257	1.13	1027	1.08	7777	0.88	1025	0.42	7777	1.27
Top10	46.15	Top10	72.42	Top10	38.15	Top10	86.18	Top10	8.64	Top10	11.16	Top10	45.22	Top10	27.58	Top10	31.60	Top10	39.96	Top10	57.51	Top10	60.39
Top20	46.74	Top20	79.54	Top20	46.33	Top20	91.18	Top20	9.16	Top20	13.98	Top20	49.24	Top20	33.06	Top20	39.90	Top20	47.13	Top20	59.09	Top20	69.93

Volume ratio varied between 0.02 and 0.11, showing that the TCP volume contributed the most traffic. The UDP volume contributed about 1% to 9%, marginally small compared to TCP. In particular, the NZIX-II-00 and LEIP-II-03 networks had the highest ratio (about 9% UDP percentages), but they showed quite different port number usages. For example, NZIX-II-00 had the most UDP volume on port 53 (DNS) and 123 (NTP) while LEIP-II-03 had the most p2p UDP volume – port 4672 (eD2k) and 6257 (WinMX). Considering the number of flows, the flow ratio varied between 0.04 and 2.00. AUCK networks, for example, have the ratio increased from 0.19 (1999) to 1.19 (2007), then decreased to 0.66 (2009). Over time the WITS and CAIDA networks also have the ratio increased up to 1.95 (2006) and 2.00 (2009) respectively. Other networks are similar, though not systematic. Compared with volume, it shows that UDP flows in general are more frequently observed than TCP, but are mainly smaller in bytes. There is no observed trend to longer, fatter UDP flows as we might expect from streaming.

One reason why the flow ratios might fluctuate a lot, even for the same network, is that UDP seems to be used a lot for malicious transmission. A port scan, for example, generates many flows containing only a single packet by enumerating a large range of port numbers. Another reason might likely to be due to small-sized signaling flows, which are often used by emerging applications.

Table II shows six selected network's top10 most used port numbers, ranked according to their proportions for flows, volume and duration. It also shows a cumulated percentage of these top10 and top20 ports. Figure 3 and Figure 4 shows the cumulative distribution function (CDF) plot – the top two plots are for TCP, showing port numbers on a linear and a log scale respectively, and the bottom two plots are for UDP. Due to space constraints, only selected networks are shown, and the rest of the tables and plots are shown in [19].

Overall, the top10 flows together contributed about 18% (ISP-B-05) to 60% (CAIDA-DirA-09) for TCP, and 9% (CAIDA-DirB-09) to 76% (SITE-I-03) for UDP. The ranges for the top10 volumes were greater, i.e., 33% (ISP-B-05) to 88% (AUCK-09) for TCP, and 11% (CAIDA-DirB-09) to 86% (BELL-I-02) for UDP. We find little systematic trend for both TCP and UDP; these variabilities show that the traffic can either be heavily dominated by a few port numbers, or diversely dispersed. Various other well-known port numbers (up to 1023) also contributed to the top10. The individual port usages are less significantly contributed for higher ranks, e.g., top20 increased pecentages only slightly.

For TCP, we observe that HTTP/S (80/443) traffic con-

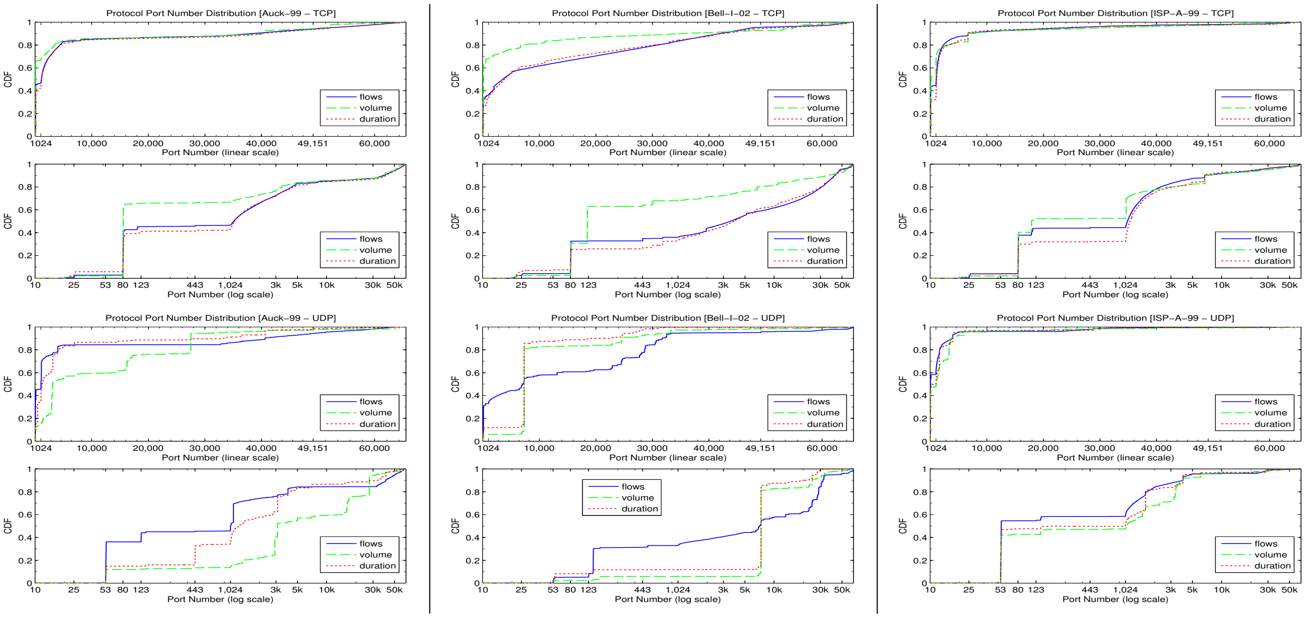

Fig. 3. Port Number Distribution – Older networks, Left:`AUCK-99`, Center:`BELL-I-02`, Right:`ISP-A-99`

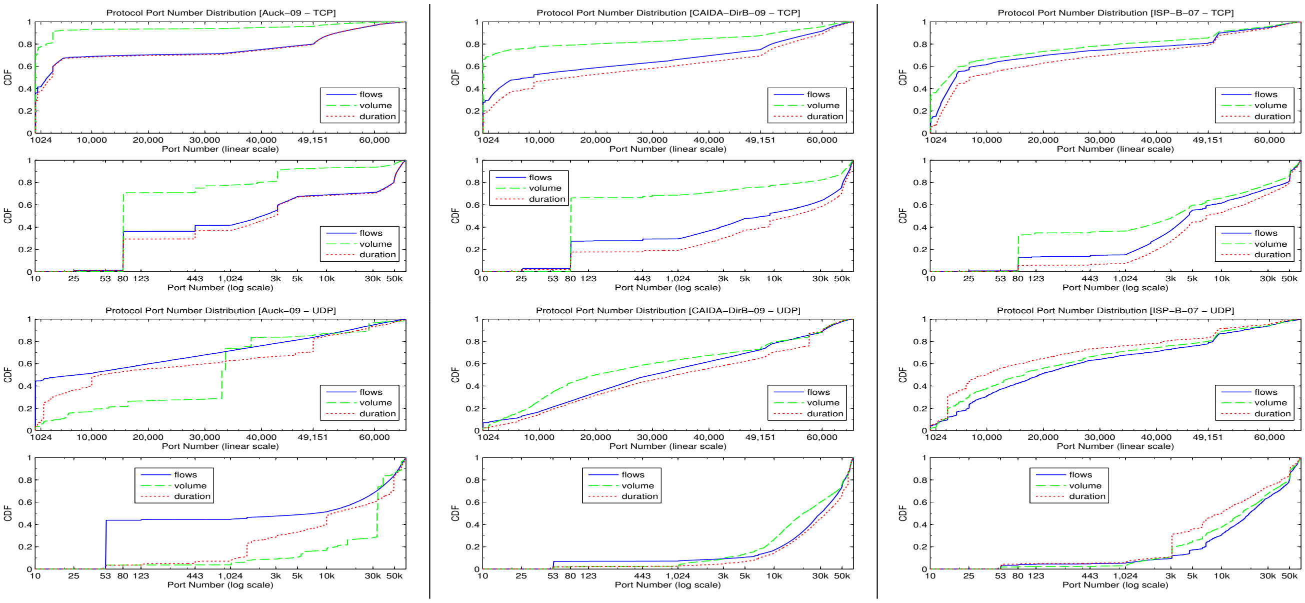

Fig. 4. Port Number Distribution – Newer networks, Left:`AUCK-09`, Center:`CAIDA-DirB-09`, Right:`ISP-B-07`

tributed the most and often appeared in the top rank. We also observe that generally recent networks have more high-end port numbers compared to the older networks. For UDP, DNS traffic were the most common, although rank distributions appear similar between the networks, we observe that the distributions are less skewed over the years, given that their volumes are already marginally small. Volumes on the port numbers are more diversely spread over the years, e.g., top10 volumes have reduced from 77% to 53% (`WITS-04` to `WITS-06`), and only less than 17% of UDP volumes (`CAIDA-DirA-09`, `CAIDA-DirB-09`, `ISP-B-07`) are observed. These changes show that there are more applications using different port numbers in recent years. None of these ports however indicate any plausible evidence of incremental streaming traffic.

We observe how the port numbers are distributed by their attributes – number of flows and volume/duration. Measuring the volume for a particular port number is the same as measuring an aggregated flow size on that port number. Similarly, duration measures the total aggregated flow lifetimes of a given port number. Here, we find that often up to 70% to 90% of port numbers used are below 10,000. The rest of the port usage appears quite uniformly distributed, although not strictly linear. A step in the CDF for one particular port number shows that this port is heavily used in the network being studied, e.g., FTP/SMTP and HTTP/S traffic, which is to be expected for well-known ports or registered ports.

The registered ports are those from 1024 to 49151, so steps in the CDF are to be expected throughout this range. We do see this in several plots, for both UDP and TCP. We also see a roughly linear CDF for ports in the dynamic range above 49151, which is to be expected if they are chosen pseudo-randomly, as good security practice requires. The situation between 1024 and 49151 is somewhat confused, because many TCP/IP implementations appear to use arbitrary ranges between 1024 and 65535 for dynamic ports (often referred to as "ephemeral" ports, which is not a term defined in the TCP or UDP standards or in the IANA port allocations). It appears that different Operating Systems, as well as their different versions, use a different range by default [9].

Both volume and duration distributions appear similar to the flow distribution, i.e., increase in the number of flows also increases total volume and durations. Some port numbers do not correlate equally with flows, volume and duration. For example, BELL-I-02 contained almost no flows on port 7331, but those flows carried more than 70% of volume and duration. Similarly, SITE-I-03 contained 0.4% of FTP data flows, but those contributed more than 43% of volume.

For older traces, a majority of protocols are low numbered, e.g., ISP-A-99 have more than 90% of traffic flows and volumes contributed to port number below 10,000, for both TCP and UDP. Conversely, recent traces have only up to about 50% (ISP-B-07). UDP traffic is a lot more linearly distributed across the port range, e.g., both CAIDA-DirB-09 and ISP-B-07. Also, DNS traffic volumes are no longer significant, e.g., contributing from 42% (ISP-A-99) to less than 2% (ISP-B-07). These changes appear to be the major differences between the older and newer traces, given that the volume ratios hardly changed.

IV. Discussion

The UDP to TCP ratio does not seem to show any systematic trend; there are variations over time and between networks, but nothing we can identify as characteristic. In particular, there is nothing in the data to suggest a sustained growth in the share of UDP traffic caused by growth in audio and video streaming. Although we have observed a diversity of port numbers increasing over time, recent (2009) traffic volume appears to be aggregated on HTTP/S, and thus a prediction of increasing web traffic could be reasonable (e.g., [5]). It appears that a large number of application developers are taking advantage of and utilizing web traffic to increase inter-operability through NATs and firewalls, mitigating deployment and operation issues [18]. From this, we may again observe the top port ranks contributing a lot more HTTP/S traffic, making the volume distributions similar to older network traffic.

It also appears that DNS traffic that was once a main contributor of UDP volume no longer stands out; instead UDP port numbers are more spread, presumably due to application diversities, possibly including streaming traffic. In fact, superficial evidence suggests that popular streaming solutions are at least as likely to use TCP (with or without HTTP) as they are to use UDP (with or without RTP). Our observations cannot directly detect this, but it is certain that we are not seeing a significant shift from TCP to UDP. Since streaming traffic is believed to be increasing, we must have an increase in the amount of TCP traffic for which TCP's response to congestion and loss (slowing down and retransmitting) is counter-productive.

In many cases, there are correlations of our three attributes, e.g., port 80 with a high proportion of flows is also likely to have a high proportion of both volume and duration. Similarly, an unpopular port number is likely to have low values for flows, volume and duration. However, certain ports with a low number of flows could contribute a high volume of traffic. Port usage trends are obviously dependent on application trends. As we have seen, these vary between networks, so local observations are the only valid guide. This could be significant if a service provider is planning to use any kind of address sharing by restricting the port range per subscriber [21]. There seems to be no general rule about which ports are popular, except for the few very well known service ports.

Our observations of port usage also shows considerable but not systematic variation between networks. This is somewhat surprising; all the networks are large enough that we would expect usage patterns to average out and be similar in all cases. We can speculate that the demographics of the various user populations (e.g., students and academics versus general population) cause them to use rather different sets of operating systems and applications. However, the main lesson is that one cannot extrapolate from usage patterns on one network to those on another without allowing for at least as much variability as we have observed in this study.

From this, our observations also suggest several guidelines for potential measurements on operational networks. First, variation in the number of flows may indicate network instabilities and abnormal behaviors. The observed variability implies that one needs to be flexible when configuring the measurement parameters, e.g., the traffic meter's flow table size, perhaps adjusting the flow timeout differently for each port number. Second, the volume and duration of flows indicate potential network improvements based on port usages; in the port and rank distribution, the slopes indicate how the port numbers are concentrated in small or large ranges. These information can be considered for purposes such as prioritizing specific applications of interest, or new strategy in load balancing and accounting/billing. Flow-based routing (for example, [22]) has the ability to resolve integrity of inelastic traffic by keeping track of flows for faster routing, though little evidence of applications has been reported.

V. Related Work

We note that port-based observations can give inaccurate protocol identification; however studies have shown (e.g., [17], [18]) that port numbers still give reasonable insights into applications and trends. Faber [12] suggested that IP hosts producing UDP flows could be characterized by weight functions, e.g., between p2p and scans. Also, McNutt and De Shon [20] have computed correlations in the usage of ephemeral ports to identify potential malicious traffic patterns. Wang *et al.* [24] reported on a short term study of the distribution of ephemeral port usage; they consider any port above 1024 to be ephemeral, not distinguishing between the registered and dynamic ports. Ephemeral port number cycling can be visualized so as to detect hidden services [13]. Allman [11] suggested different ways to select ephemeral ports that are more diverse and robust against security. Much interest in the choice of ephemeral port numbers was aroused by the DNS vulnerability publicized in 2008 [3]. It is to be expected that as developers learn the lesson of this vulnerability, randomization of port numbers may become more prevalent.

VI. Conclusion

In this report, we have have observed two widely used protocols (UDP and TCP) to measure how their $\frac{UDP}{TCP}$ ratio varied. Particularly we observed that there is no clear evidence that the ratio is increasing or decreasing. The ratio is rather dependent on application popularity and, consequently, on user choices. The volume ratio had subtle variations – the majority of volume is dominated by TCP, with a diurnal pattern. The flow ratio had larger variations – many flows are UDP but with very small volume.

Although the ratio does not vary systematically among the networks, each had quite different port number distributions. For example, data from recent years of ISP networks contained a significant amount of p2p traffic, while enterprise networks contained a large amount of FTP traffic. Again, user choices are at work. There were however no particular signs of incremental use of well-known port numbers for audio or video streaming.

As we note that emerging applications use arbitrary port numbers, identifying applications solely based on port numbers alone could lead to inaccurate assumption; deep packet inspection may be the only approach in practice to determine the streaming traffic, provided that the packets are not encrypted. It could continue to be, on the other hand, that the streaming concepts may simply further be evolved or integrated into elastic data traffic, provided that the over-provisioning is considerably tolerated. Nevertheless, the trend towards more streaming traffic seems undeniable. However, contrary to what might naively be expected, there is no evidence of a resulting trend to relatively more use of UDP to carry it. In fact, the evidence is of widespread variability in the fraction of UDP traffic. Similarly, there is no clear trend in port usage, only evidence of widespread variability.

We had hoped to derive some general guidelines about the likely trend in traffic patterns, particularly concerning the fraction of non-congestion-controlled flows and the distribution of port usage. There appear to be no such guidelines in the available data. We consider that router and switch designers, as well as network operators, should be well aware of high variability in these basic characteristics, and design and provision their systems accordingly. In particular, one cannot extrapolate from measurements of one user population to the likely traffic patterns of another. It seems that all network operators need to measure their own protocol and port usage profiles.

Acknowledgments

Preliminary data on UDP to TCP ratios was kindly supplied by Arnold Nipper, Toshinori Ishii, Kjetil Olsen, Mike Hughes and Arne Oslebo. We are grateful to Ryan Koga of CAIDA and to Stanislav Shalunov, formerly of Internet2, for information about their respective datasets. The work reported here was partially supported by Huawei Technologies Co. Ltd.

References

[1] "Analyzing UDP usage in Internet traffic," http://www.caida.org/research/traffic-analysis/tcpudpratio/.

[2] "CAIDA Internet Data – Realtime Monitors," http://www.caida.org/data/realtime/index.xml.

[3] "CERT Vulnerability Note VU#800113," http://www.kb.cert.org/vuls/id/800113/.

[4] "Cisco Visual Networking Index: Usage Study," http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/Cisco_VNI_Usage_WP.pdf.

[5] "Hyperconnectivity and the Approaching Zettabyte Era," http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.pdf.

[6] "Internet2 NetFlow: Weekly Reports," http://netflow.internet2.edu/weekly/.

[7] "Minnesota Internet Traffic Studies (MINTS)," http://www.dtc.umn.edu/mints/home.php.

[8] "Passive Measurement and Analysis (PMA)," http://pma.nlanr.net/.

[9] "The Ephemeral Port Range," http://www.ncftp.com/ncftpd/doc/misc/ephemeral_ports.html.

[10] "WITS: Waikato Internet Traffic Storage," http://www.wand.net.nz/wits/.

[11] M. Allman, "Comments on selecting ephemeral ports," *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 2, pp. 13–19, 2009.

[12] S. Faber, "Is there any value in bulk network traces?" *FloCon*, 2009.

[13] J. Janies, "Existence plots: A low-resolution time series for port behavior analysis," in *VizSec '08: Proceedings of the 5th international workshop on Visualization for Computer Security*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 161–168.

[14] Joe St Sauver, University of Oregon, "Personal communication," 2008.

[15] W. John and S. Tafvelin, "Analysis of internet backbone traffic and header anomalies observed," in *IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. New York, NY, USA: ACM, 2007, pp. 111–116.

[16] T. Karagiannis, A. Broido, N. Brownlee, K. Claffy, and M. Faloutsos, "Is p2p dying or just hiding?" in *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, vol. 3, Nov.-3 Dec. 2004, pp. 1532–1538 Vol.3.

[17] H. Kim, K. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K. Lee, "Internet traffic classification demystified: myths, caveats, and the best practices," in *CONEXT '08: Proceedings of the 2008 ACM CoNEXT Conference*. New York, NY, USA: ACM, 2008, pp. 1–12.

[18] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, F. Jahanian, and M. Karir, "2009 Internet Observatory Report," http://www.nanog.org/meetings/nanog47/presentations/Monday/Labovitz_ObserveReport_N47_Mon.pdf, 2009.

[19] D. Lee, B. Carpenter, and N. Brownlee, "Observations of UDP to TCP Ratio and Port Numbers, (Technical Report)," http://www.cs.auckland.ac.nz/~brian/udptcp-ratio-TechReport.pdf, 2009.

[20] J. McNutt and M. D. Shon, "Correlations between quiescent ports in network flows," *FloCon*, 2005.

[21] R. Bush (ed.), "The A+P Approach to the IPv4 Address Shortage (work in progress)," http://tools.ietf.org/id/draft-ymbk-aplusp, 2009.

[22] L. Roberts, "A radical new router," *Spectrum, IEEE*, vol. 46, no. 7, pp. 34–39, July 2009.

[23] J. Rosenberg, "UDP and TCP as the New Waist of the Internet Hourglass," http://tools.ietf.org/id/draft-rosenberg-internet-waist-hourglass-00.txt.

[24] H. Wang, R. Zhou, and Y. He, "An Information Acquisition Method Based on NetFlow for Network Situation Awareness," *Advanced Software Engineering and Its Applications*, pp. 23–26, 2008.