

A Flow-Based Performance Analysis of TCP and TCP Applications

Lei Qian

Department of Computer Science
University of Auckland
Auckland, New Zealand
Email: lqia012@aucklanduni.ac.nz

Brian E. Carpenter

Department of Computer Science
University of Auckland
Auckland, New Zealand
Email: brian@cs.auckland.ac.nz

Abstract - The most widely used transport layer protocol, TCP, consistently contributes the largest traffic volume in the Internet. The objective of this study is to determine whether there are long-term trends in TCP flows and popular TCP applications from over several years in different organizations. The results show that TCP flows' average packet size has increased steadily and each TCP flow contains slightly more packets year by year. HTTP packets have shorter Round Trip Time (RTT) than HTTPS packets. We also observed that over 60% of TCP flows have very short average packet length. Most TCP flows have reasonably short lifetimes.

Keywords-TCP;TCP applications;long-term impacts

I. INTRODUCTION

Internet usage has increased explosively over recent decades. With continuous increase in the number of computers connected to the Internet [6], more and more applications, particularly for multimedia, are now using the Internet. A common prediction is that the proportion of UDP traffic will increase steadily because many streaming protocols, such as RTP, were originally designed for UDP. However, according to a recent report [1], there is no evidence that the percentage of UDP traffic has been increasing over the years. A number of studies have found that over 85% of total Internet traffic is TCP traffic [1, 2]. Based on the port number analysis in [1], well known ports 80 (HTTP) and 443 (HTTPS) are the most popular both for volume and flow counts. Ports 21 (FTP) and 25 (SMTP) have also been in the top 10 port usage list over the years. However, application layer protocols over TCP typically generate most traffic volume. Studies [8, 9] proposed streaming protocols over TCP or HTTP, and streaming over HTTP seems to have become an acceptable solution in practice recently.

The aim of this work is to understand whether there are systematic trends over recent years in TCP flows, and whether the most popular application layer protocols used over TCP show any particular long-term. The five application layer protocols that we considered are FTP, HTTP, HTTPS, SMTP and SSH, referred to below as the "five applications".

II. OVERVIEW OF DATA AND MEASUREMENT METRICS

A. Data Sets Overview

Table I shows basic information on our measurement trace files, which includes date, time, duration, and volume and packet count. We use eight existing data traces from The University of Auckland and The University of Waikato Internet Traffic Storage (WITS) [3] from the WAND group [4]. Over seven years, more than 3 TB of traffic and over 4.7 billion packets have been measured on the two campuses.

B. Measurement Metrics

A TCP flow can be identified as a group of packets with the same five attributes: source IP, destination IP, source port, destination port and protocol number [5]. For this study, we decided that during the analysis of a given trace, a TCP flow would be considered to end either on TCP FIN/ACK, or if the hash table used in the analysis reached its size limit, or if a given flow was silent for a fixed timeout of 30 seconds, to prevent our analysis program running out of memory [1]. We have focused on four important parameters - Flow packet size, Flow Duration, Number of Packets and Round Trip Time. We removed Flow volume from consideration because we observed that over 95% of TCP flows have less than 3 MB of data. Parameters are defined as follows:

1. *Flow's Average Packet Size (bytes)*: the total packet volume in a single two-way flow divided by total packet count, i.e., the average information carried within a packet in a single flow, including packet headers and acknowledgements.
2. *Flow's Duration (seconds)*: the flow lifetime, i.e., the time difference between the last packet and the first packet seen within an individual flow. Flow Duration has been considered as a useful measurement metric for traffic prediction or accounting [5].
3. *Flow's Packet Count*: Total number of packets observed in each flow.
4. *Packet Round Trip Time (RTT) (ms)*: RTT has been defined by IETF in [7] to measure single packet delay. In practice, we are able to use the timestamp information tagged on each packet and use the corresponding TCP ACK to determine a packet's RTT value.

TABLE I. SUMMARY OF TRACE FILES

Trace File	Start Time	Duration	Volume (Gigabytes)	Number of Packets
------------	------------	----------	--------------------	-------------------

AKL-2003	2003-Dec-04 [13:42]	24.00hr	68.23	150237043
AKL-2005	2005-Aug-16 [14:00]	03.00hr	48.42	89092876
AKL-2006	2006-July-27 [13:00]	24.00hr	294.30	486953987
AKL-2007	2007-Nov-01 [13:00]	24.00hr	652.41	1077489105
AKL-2009	2009-Aug-03 [08:00]	11.00hr	1860.74	2534849069
WITS-2004	2004-Mar-01 [00:00]	24.00hr	37.29	91425148
WITS-2005	2005-May-12 [00:00]	24.00hr	58.40	138829702
WITS-2006	2006-Oct-30 [00:00]	24.00hr	79.25	173121901

TABLE II. USAGE OF FIVE TCP APPLICATIONS

	AKL-03		AKL-05		AKL-06		AKL-07		AKL-09		WITS-04		WITS-05		WITS-06	
	Vol. %	Flow %	Vol. %	Flow %	Vol. %	Flow %	Vol. %	Flow %	Vol. %	Flow %	Vol. %	Flow %	Vol. %	Flow %	Vol. %	Flow %
FTP	5.35	1.735 E-2	3.25	4.30 E-2	2.14	6.48 E-2	1.24	3.69 E-3	1.67	5.30 E-3	4.46	6.34 E-3	1.79	2.78 E-2	1.80	1.67 E-3
HTTP	64.33	31.76	73.97	69.48	52.66	55.48	54.08	57.62	75.35	66.54	63.85	46.53	71.28	49.24	67.69	55.22
HTTPS	8.46	5.82	5.02	6.16	5.71	9.20	4.83	13.19	5.06	11.41	12.26	7.98	8.73	18.85	11.29	9.94
SMTP	6.44	6.70	3.50	1.50	2.82	5.39	2.77	11.60	1.55	1.87	6.69	4.53	8.43	8.99	8.77	14.83
SSH	6.35	5.78 E-2	5.01	7.71 E-2	4.72	1.68	1.29	8.65 E-2	0.25	0.30	4.03	5.43 E-2	2.60	0.74	0.50	0.69

III. RESULTS

In this section, we present and interpret our analysis results. Tables and Cumulative Distribution Function (CDF) plots are used to seek long-term trends, based on the eight dated trace files and the five applications.

Table II shows the five applications' relative contribution to volume and flow count. At least half of TCP traffic is HTTP traffic. We found SSH and FTP contributed proportionately less TCP traffic in recent years (**AKL-2007**, **AKL-2009** and **WITS 2005**, **WITS 2006**) than in previous years (**AKL-2003**, **AKL-2005**, **AKL-2006** and **WITS 2004**), but little systematic trend is shown in the table. The percentage of HTTPS and SMTP traffic is variable over the years. The proportion of SSH traffic noticeably decreases over the years.

A. TCP Results by years.

Figure 1 shows the CDF plot of average packet size of each TCP flow. Surprisingly, many flows with small average packets were found. Both **AKL** and **WITS** plots indicate that average packet size tends to increase over the years. **AKL-2003** shows that over 50% of TCP flows have an average packet size of only about 80 bytes; **WITS-2004** similarly shows that about 50% of TCP flows have this average packet size. The recent plot **AKL-2009** shows only about 25% of flows with average packet size of 80 bytes and **WITS-2006** has even fewer. This trend might indicate systematic change in network conditions or in TCP applications. In the early years, large proportions of TCP traffic were simple HTTP GET and POST requests. This could explain the many flows with short

average flow length shown in Figure 1. A possible reason might be that more and more applications other than HTTP GET and POST are appearing over TCP, causing a significant increase of flow average packet size.

Figure 2 shows the CDF plot of Packet Count for each TCP flow. Overall, although we observed that there are always several very long TCP flows (with the same duration as the capture file itself) in each trace, approximately 90% of TCP flows have no more than 50 packets, with slightly more packets sent per individual TCP flow over the years. In earlier years, over 50% of flows only contain two to five packets in a flow for both **AKL-2003** and **WITS-2004**. The **AKL-2009** and **WITS-2006** show that up to 50% of TCP flows have no more than 25 packets in each TCP flow.

Figure 3 shows the CDF plot of Flow Duration is getting very slightly longer over the years. **AKL-2003** to **AKL-2009** does not show any systematic trend, but **WITS-2006** flow duration is slightly longer than **WITS-2004** and **WITS-2005**. All curves show that about 90% of TCP flows have lifetimes below one minute.

The results illustrated in Figure 2 and Figure 3 may be biased by the parameters of our measurement software, for example, the 30 seconds of fixed timeout value and the internal hash table size. Nevertheless, the most interesting findings were the short duration flows and small packet counts in each flow. One factor may be that all traces analyzed in this report are gathered from educational organizations. Other types of networks, such as ISP networks, might have significantly different results.

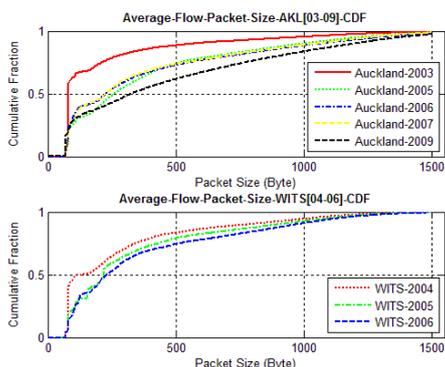


Figure 1. TCP Flows Average Packet Length CDF

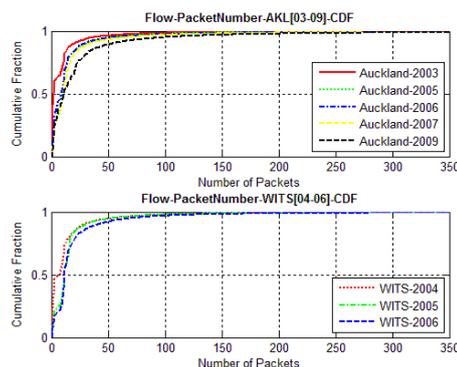


Figure 2. TCP Flows Packet Count CDF

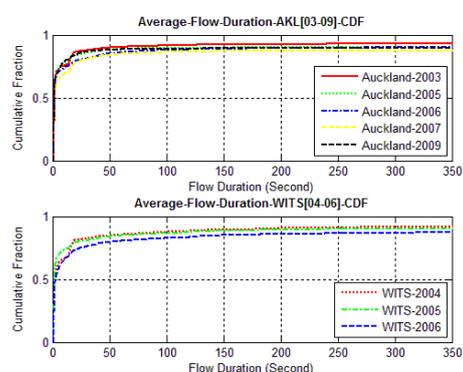


Figure 3. TCP Flows Duration CDF.

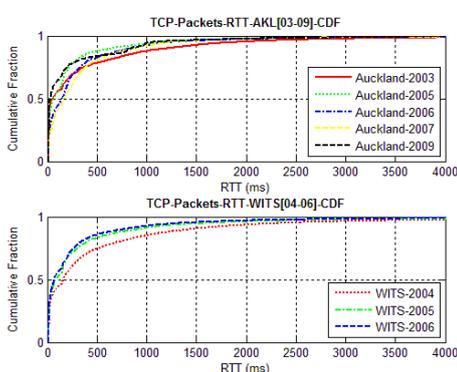


Figure 4. TCP Packets RTT CDF

Figure 4 shows the CDF plot of observed TCP RTT. Over 75% of TCP packets are acknowledged within 500ms. There is no evidence of any long term trend in the AKL [03-09] trace files. Slightly larger fractions of TCP packets have been acknowledged within 500ms in AKL-2005 compared to AKL-2009. The WITS [04-06] plot shows that distinctly more TCP packets are acknowledged within the same RTT over the years. The difference between WITS and AKL could have many explanations, such as a relative decrease in congestion in the Waikato campus network.

B. TCP five applications result

Figure 5 to Figure 8 illustrate the results in more detail for the TCP five applications. Fig. 5, 6 and 7 show Flow Average Packet length, Flow count and Flow Duration CDF distribution, and Fig. 8 shows packet RTT CDF distributions of the five applications across the years.

Figure 5 shows that SMTP and SSH have the two shortest average packet lengths in each flow. A possible explanation might be SSH was originally designed for sending short packets with interactive response needed, but the small average size observed for SMTP is quite surprising. Surprisingly, FTP does not always have the longest packet length; AKL-2005 shows more than 90% of FTP packets with no longer than 250 byte length, and AKL-2006 shows that FTP has a similar distribution to HTTPS. FTP was of course designed for file transfer, so a longer average packet length

would be expected. It may be that both SMTP and FTP traffic includes a high proportion of control messages compared to user data messages. FTP average flow packet lengths are variable over both years and organizations. HTTP and HTTPS have similar distribution in AKL-2003, WITS-2004 and WITS-2006. Most plots shows that HTTPS has shorter average flow packet length than HTTP (AKL-2005, AKL-2006, AKL-2007 AKL-2009 and WITS 2005).

Figure 6 shows the packet count CDF distribution of five applications. Over about 80% of FTP, HTTP/S and SSH flows have no more than 50 packets per flow. FTP distributions are variable. FTP distributions in AKL-2003, AKL-2005, AKL-2006, AKL-2009 AND WITS-2005 are similar to HTTP/S. AKL-2007 shows that about 20% of FTP flows have more than 350 packets and WITS-2006 was similar.

Figure 7 shows the five applications' duration for each flow. Our expectation was that SSH would have the longest duration with small volume because SSH is commonly used for remote login. The characteristic of SSH in principle was lasting long time, transferring less volume with fast response [10]. Contrary to expectations, Figure 7 does not illustrate a significant difference between SSH and the other four applications, except for WITS-2004. Almost all five application flows are terminated within 350 seconds.

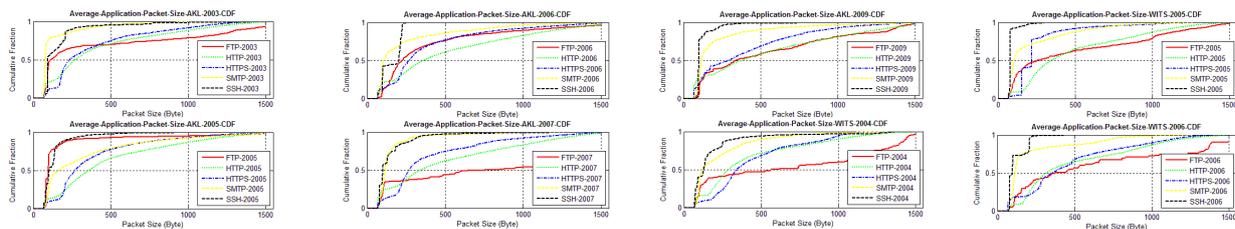


Figure 5. Five applications Flow Average Packet Length CDF

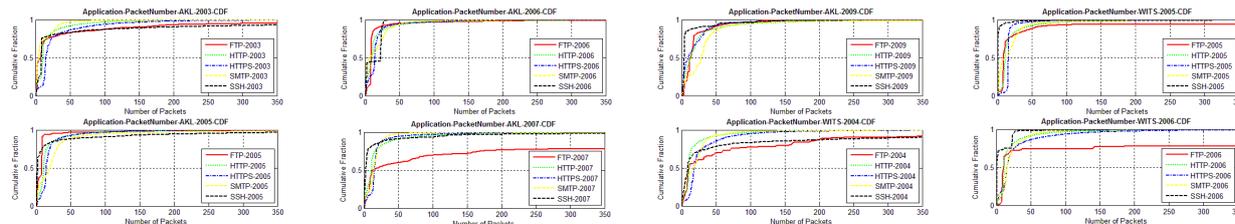


Figure 6. Five applications Flow Packet Count CDF

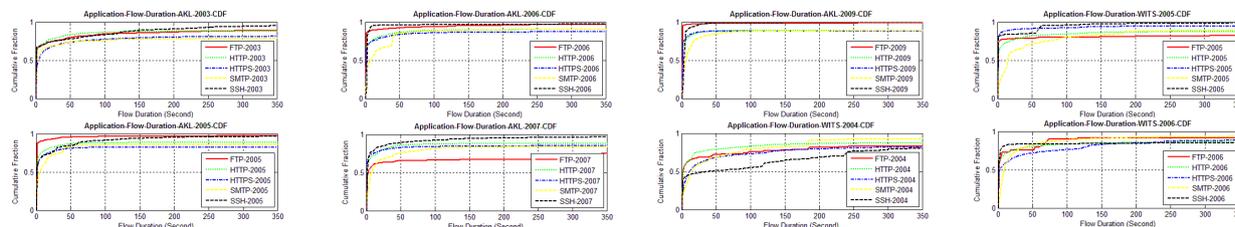


Figure 7. Five applications Flow Duration CDF

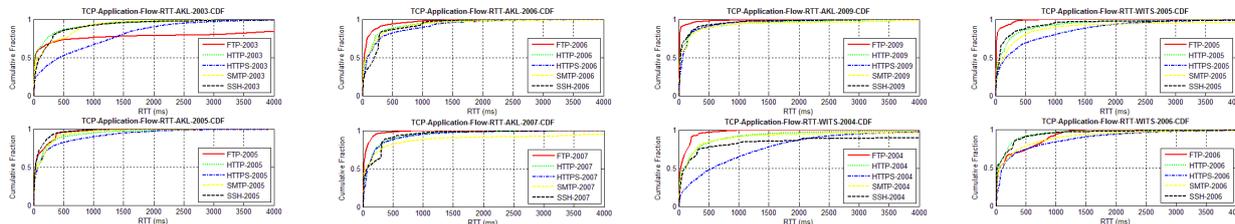


Figure 8. Five applications Packets RTT CDF

A possible reason for this might be our fixed-timeout value or the size of hash table. We might inadvertently split a long SSH flow into several shorter flows.

Figure 8 shows the RTT CDF distribution for the five applications. SSH traffic is acknowledged within a reasonably short time; about 80% of SSH packets are acknowledged within half a second. We observed that HTTP traffic may have shorter response time (AKL-2003 to AKL-2009 and WITS-2004 to WITS-2006), and HTTPS is slower than both. This may be because SSH and HTTPS both involve cryptographic calculations. FTP's RTT is usually smallest (in AKL-2006, AKL-2007, AKL-2009, WITS-2004 and WITS-2005).

Overall, there is no clear evidence of any long term trend in our results, which does not match our expectations. The results only illustrate the data we analyzed from two particular networks, and could be affected by hardware performance, traffic load, time of the day/month and possibly other issues.

IV. CONCLUSION

In this report, we have presented an analysis of TCP and five TCP applications, based on average packet size, packet counts, flow duration, and RTT for each trace file. The findings show that flows' average packet length has increased over years and HTTP experiences shorter RTT than HTTPS. No clear evidence shows that application-specific protocols like FTP or SMTP have any particular attributes compared to more generic protocols like HTTP.

We note that some traffic, such as long-duration, low-volume SSH sessions, may be misrepresented in our results. Resolving this would require supplementary analysis tools. Also, more studies to compare networks in different years and different places would be useful. However, we consider this report is of value to network operators who design, implement and manage networks, especially for educational campus networks.

ACKNOWLEDGMENT

We are grateful to Associate Professor Nevil Brownlee (The University of Auckland) for his constructive advice and valuable comments. This report would not have been possible without his advice and support. We are also grateful to Dr Dongjin Lee (The University of Auckland) for various network measurement discussions. We thank Richard Nelson and the WAND research group (The University of Waikato) for providing us resources and useful information about network trace files.

REFERENCES

- [1] D. Lee, B. Carpenter and N. Brownlee, "Observations of UDP to TCP Ratio and Port Numbers", Fifth International Conference on Internet Monitoring and Protection (ICIMP 2010), Barcelona, May 2010.
- [2] R. Nelson, D. Lawson and P. Lorier, "Analysis of Long Duration Traces", ACM SIGCOMM Computer Communication Review, Volume 35 Issue 1, January 2005.
- [3] Waikato Internet Trace Storage.
<http://wand.cs.waikato.ac.nz/wand/wits.inex.html>
- [4] WAND Network Reserch Group
<http://wand.cs.waikato.ac.nz>.
- [5] D Lee and N.Brownlee, "Passive Measurement of One-way and Two-way Flow Lifetimes", ACM SIGGCOM Computer Communication Review, Volume 37 Issue 3, July 2007.
- [6] B. Carpenter, "Observed Relationships between Size Measures of the Internet", ACM SIGGCOM Computer Communication Review, Volume 39 Issue 2, 6-12, April 2009.
- [7] G. Almes, S. Kalidindi and M. Zekauskas, "A Round-trip Delay Metric for IPPM", Internet RFC 2681, September 1999.
- [8] B. Mukherjee and T. Brecht, "Time-Lined TCP for the TCP Friendly Delivery of Streaming Media", Proc. IEEE Int. Conf. Network Protocols, Osaka, Japan, November 2000.
- [9] Puneet Mehra, "Efficient Video Streaming over TCP", available at: <http://www.eecs.berkeley.edu/IPRO/Summary/Old.summaries/03abstracts/pmehra.1.html>
- [10] B. Carpenter and K. Nichols, "Differentiated Service in the Internet", Proc. IEEE , Page 1479-1494, September, 2002.