Advances in Computer Science Research, volume 88
International Conference on Computer, Network, Communication and Information Systems (CNCI 2019)

Analysis of TCP Flow Characteristics

Xiaoyu Cui, Xingchen Xu, Qingguo Shen and Xuefei Liang
The Army Engineering University of PLA Nanjing, China
axiaoyu940528@sina.cn, btammyxxc@163.com, cshenqg2@163.com,18362973763@163.com

Abstract. The coexistence of long-lived TCP flow and short-lived TCP flow is studied through the analysis of a large number of literatures. Extract long-lived TCP flow and short-lived TCP flow and count the traffic characteristics of TCP flow, including the flow number, maximum flow rate, average rate and average duration of long-lived TCP flow and short-lived TCP flow. It also analyzes the flow characteristic rule and the relationship between various flows, draws the graph line, in order to present the network performance more intuitively.

Keywords: Traffic analysis, TCP, long-lived TCP flow, short-lived TCP flow.

1. Introduction

In recent years, with the rapid development of information technology and network technology, network has become an indispensable part of daily work and life, playing an increasingly important role, and has become a basic requirement for national development and social progress. Traffic characteristic analysis and research is the basic means to understand the network, which carries extremely rich information. It has very important research value for network performance analysis, network behavior prediction, network planning and traffic control.

Generally speaking, network application service needs to use transmission protocol when transferring data. At present, almost all network flow transmission uses TCP or UDP protocol, of which about 95% is transmitted through TCP. TCP is connection-oriented and reliable. The UDP protocol is connectionless. Emails, files, websites, etc. use TCP protocol, and video, audio and other data are transmitted through UDP protocol. During the study, TCP flows are usually divided into long-lived TCP flow and short-lived TCP flow according to the duration of the flow and the size of the load data. long-lived TCP flow is characterized by large amount of data and long duration. In contrast, short-lived TCP flow is basically in a slow start state and the amount of data transmitted is small. Many experimental results show that when long-lived TCP flow and short-lived TCP flow coexist in the link, they will adversely affect each other.

2. Related Work

TCP always occupies the main body of network traffic. In 1997, Thompson et al. first pointed out that TCP flow dominated in the analysis and research report of traffic characteristics of commercial backbone [1].McReary and Chaffy published the CAIDA report in 2000, pointing out that the percentage of TCP streams and UDP streams remained basically stable in the 10-month traffic data collected on NASA Ames Internet Exchange link [2].DongJin Lee and others for during the period of 1998 to 2008, commercial and academic network of various measurement data has carried on the preliminary investigation, the results showed that the UDP and TCP measured by the number of packets to the ratio of between 5% and 20% change almost no significant fluctuation, UDP flow number although increased bearing of the relatively small amount of data, shows the TCP flow to occupy a dominant position [3].The reason for such proportion of network traffic is that although the number of network users increases rapidly, the main network behavior is only web browsing, and this traffic data is transmitted through TCP protocol.

Fig. 1 shows the data traffic characteristics of seven different data centers [4]. Through the analysis of TCP flow data volume, it is found that nearly 80% of TCP flow data volume is less than 10K bytes.

At the same time, many studies show that a small proportion of traffic carries a large amount of information [5].

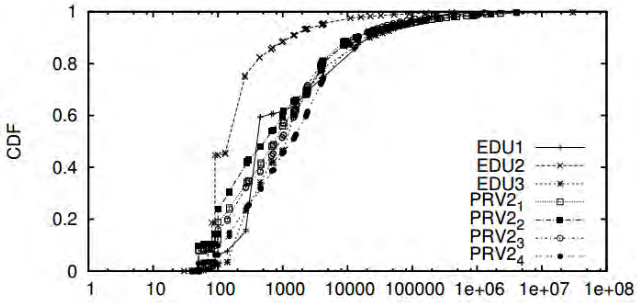


Fig. 1 TCP flow data volume distribution in different data centers

Some researchers have previously divided network traffic by different classification methods, such as data volume size ("elephant and mouse "phenomenon[6]), flow duration ("dragonflies and tortoises" phenomenon [7]), and abrupt flow (burst and worm flow [8]).In June 2002, Brownlee and Claffy conducted an eight-day traffic measurement statistics [36], and the results showed that about 1.5% of the traffic lasted more than 15 minutes, while 40% of the traffic lasted less than 2 seconds, and half of the data volume was transmitted by the flow with a longer duration.

In this article, TCP flows are divided into long-lived TCP flow and short-lived TCP flow by the duration of the flow. A stream record is defined as a packet sequence with the same quintuple, which refers to the source IP address, destination IP address, source port number, destination port number, and transport layer protocol. The continuous time distribution of TCP flows was analyzed by taking the traffic in both directions of the equinix-chicago backbone link recorded by CAIDA in 2009-2011 and 2015-2016.

Table 1 shows the characters and symbols used in this paper.

Table 1 Table of character and symbol meanings

symbol	**Meanings**
f	The number of flow
p_i	The packets of flow *i*
b_i	The bytes of flow *i*
r_i	The rate of flow *i*
s_i	The arrival time of the first packet of the flow *i*
e_i	The arrival time of the final packet of the flow *i*

3. Characteristics Analysis of TCP Flows

3.1 The Number of Flow

In 2009-2011 and 2015-2016, the flow number changes of TCP long flow, TCP short flow and abnormal flow per minute are shown in Fig. 2. As can be seen from the figure, in both directions, the number of short-lived TCP flows is far more than long-lived TCP flows, almost three or four times or more than long-lived TCP flows and the number of abnormal flows is less.

In the dirA direction, the total number of long-lived TCP flows and short-lived TCP flows is on the rise. In 2015-2016, compared with 2009-2011, the number of long-lived TCP flows increased by five or six times, which should be related to the increase of the number of users in this direction. The number of short-lived TCP flows tends to be relatively stable as the number increases.

In the dirB direction, all TCP flows fluctuate greatly, especially the short-lived TCP flows peak on several dates, reaching the maximum flow on October 29, 2010. Long-lived TCP flows are relatively stable and still account for a small proportion.

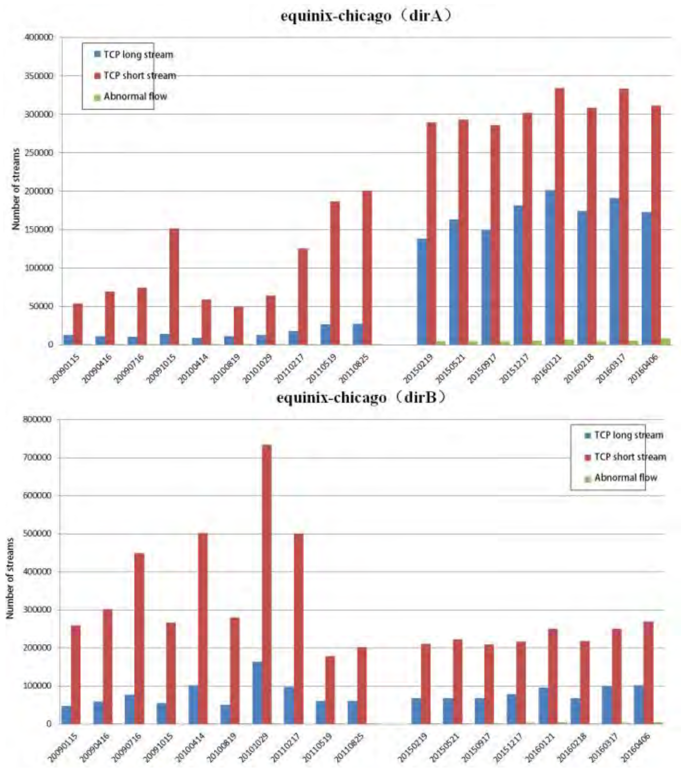


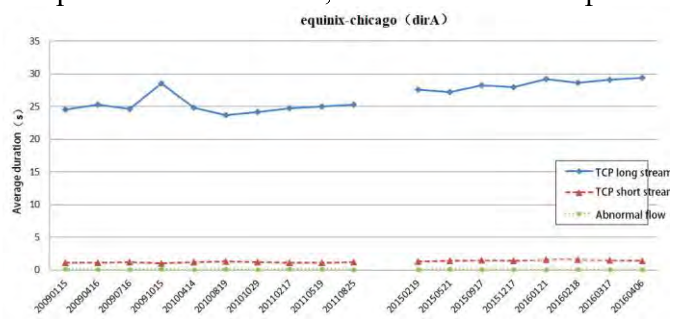
Fig. 2 The number of long-lived TCP flows and short-lived TCP flows

3.2 The Average Duration of Flow

The duration of the flow is the time interval between the first packet arrival of the flow and the last packet arrival.The average duration of the flow is defined as follows:

$$\text{avg\_duration}=\frac{\sum_{i=1}^{f}(e_i - s_i)}{f} \tag{1}$$

Fig. 3 show the average duration of long-lived TCP flows, short-lived TCP flows, and abnormal flows in the dirA direction and dirB direction. It can be seen that in 2009-2011 and 2016, the average durationof long-lived TCP flows, short-lived TCP flows and abnormal flows is basically stable in both directions, while the average duration of long-lived TCP flows is slightly increased. In the dirA direction, the average duration of long-lived TCP flows from 2009 to 2011 and from 2015 to 2016 was 25.76s, 24.24s, 25.05s, 27.78s and 29.11s, respectively. In direction B, the average duration of long-lived TCP flows was 24.38s, 24.65s, 24.82s, 27.14s and 27.33s, respectively. In these years, the average duration of short-lived TCP flows was 1.28s and the average duration of abnormal flows was 0.13s. To some extent, the duration of the flow can reflect the way users receive information. As modern people have a faster pace of work and life, information is also preferred to be simple and fast.



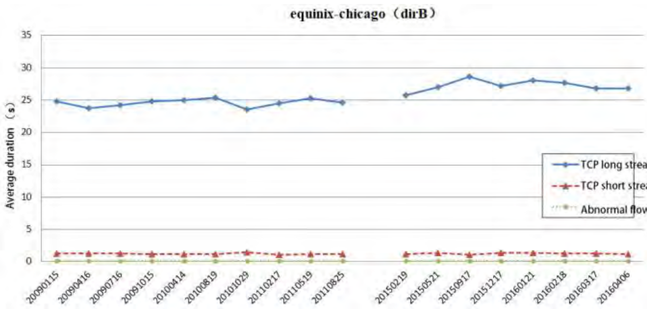


Fig. 3 The average duration of long-lived TCP flows and short-lived TCP flows

3.3 The Maximum Rate of Flow

Flow rate is the ratio of the total number of bytes of the flow to the duration of the flow, which reflects the speed of the flow.The maximum rate of traffic can be used to measure the situation of link network. When the network is prone to congestion when the traffic is too large, the rate of traffic will be limited.The maximum rate changes of long-lived TCP flows and short-lived TCP flows in two directions during 2009-2011 and 2015-2016 are shown in Fig. 4.It can be seen that the maximum rate of long-lived TCP flows is significantly greater than short-lived TCP flows, which proves that the information transmitted by long-lived TCP flows is much more than short-lived TCP flows. Since the increase in bandwidth on the equinix-chicago link in March 2015, the rate of TCP flows has grown a lot since 2015.It can also be seen from the figure that the maximum rate of long-lived TCP flows and short-lived TCP flows in 2015-2016 is sudden. In both the dirA direction and the dirB direction, the maximum rate on February 18, 2016 was smaller than that of the adjacent months, which was due to high data traffic, resulting in excessive network load and limiting the flow rate.

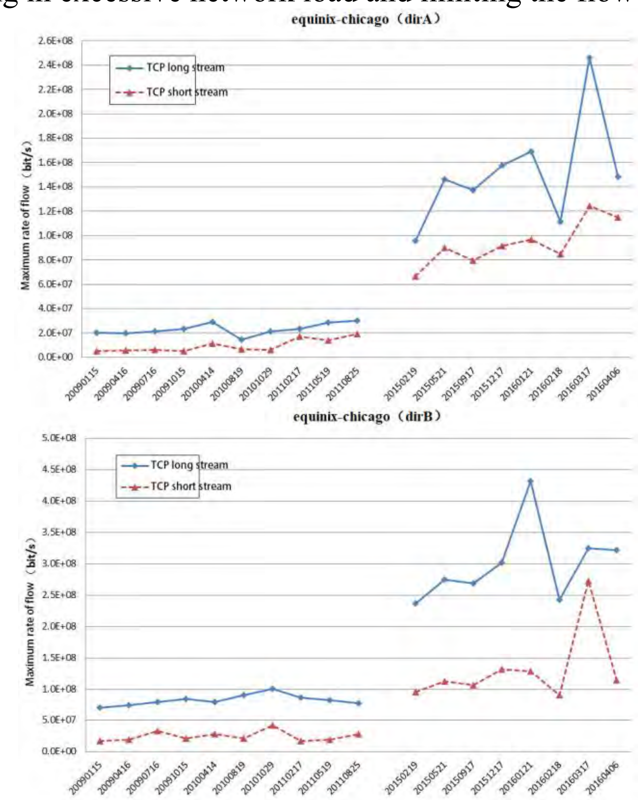


Fig. 4 The maximum rate of long-lived TCP flows and short-lived TCP flows

3.4 The Average Rate of Flow

The average rate of flow is defined as follows:

$$avg\_rate = \frac{\sum_{i=1}^{f}\left(b_i / e_i - s_i \right)}{f} \tag{2}$$

The average speed of long-lived TCP flows and short-lived TCP flows in the dirA direction and dirB direction is obtained, as shown in Fig. 5. With the average rate of the flow, we can get a general idea of the changes in link bandwidth utilization. Combined with the flow number change histogram of section 3.1 long-lived TCP flows and short-lived TCP flows, it can be confirmed that the data information of long-lived TCP flows is far more than one or two orders of magnitude of short-lived TCP flows. Meanwhile, the average speed of long-lived TCP flows is significantly faster than short-lived TCP flows. Long-lived TCP flows with a lot of information is not only a flow, but also a fast stream.

It can also be seen from the figure that the average speed of long-lived TCP flows and short-lived TCP flows over the years is also in a state of fluctuation.In dirA directions, in March 2015 after equinix - Chicago link bandwidth increases, but the average rate of long-lived TCP flows and short-lived TCP flows in 2015-2016 than that in 2009-2011 does not improve a lot, this is because the 2015-2016, long-lived TCP flows and short-lived TCP flows increased by several times, although the bandwidth increases, but also have more flow to allocate bandwidth, so the flow rate have no significant growth. In dirB directions, in October 2010, the average rate of only half of the other times, in section 3.1 we known in October 2010 the number of long-lived TCP flows increased two to three times more than the rest of the time, this also is a large number of two to three times in each flow can be assigned to the bandwidth of the lower for a third, and also significantly decreased at an average rate of flow.

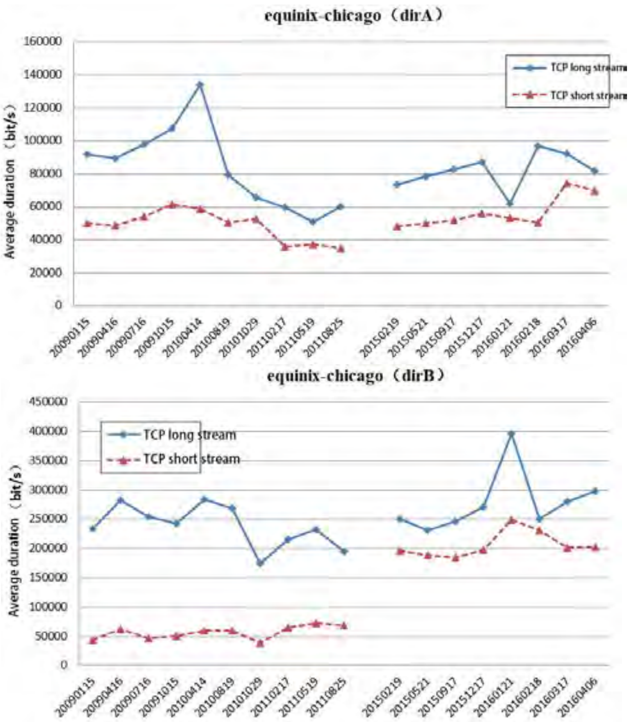


Fig. 5 The average rate of long-lived TCP flows and short-lived TCP flows

3.5 The Percentage of Bandwidth of Long-lived TCP Flows and Short-lived TCP Flows

In a network, bandwidth is used to denote the ability of the communication line of the network to transmit data.The flow number, average flow rate and average flow duration obtained from the statistics in the previous paper are obtained compute the ratio of the bandwidth of the long-lived TCP flows and short-lived TCP flows in the TCP aggregate flow in the two directions of equinix-chicago link, as shown in Fig. 5. As you can see, both the dirA direction and the dirB direction, the long-lived TCP flows have an absolutely high percentage of bandwidth. In the dirA direction, as the number of long-lived TCP flows increased significantly in 2015-2016, the percentage of bandwidth increased by at least five percentage points.In the dirB direction, long-lived TCP flows and short-lived TCP flows take a relatively stable proportion of bandwidth, and long-lived TCP flows take a ratio of 90% to 95%.

$$B = f \times avg\_duration \times avg\_rate \quad (3)$$

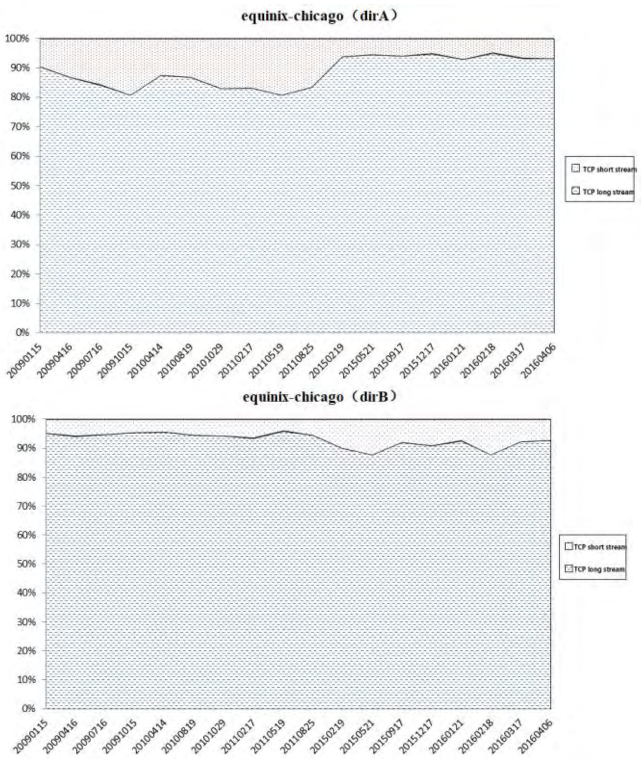


Fig. 6 The percentage of bandwidth of long-lived TCP flows and short-lived TCP flows

4. Summary

Although backbone network has the characteristics of statistical stationarity of main traffic flow, network stability will also be affected due to sudden flow. The open network measurement data sharing platform is used to collect Shared network test data in related fields for experimental research and theoretical analysis, which provides effective theoretical analysis and empirical research for network performance changes. After extracting long-lived TCP flows and short-lived TCP flows, the traffic characteristics of TCP flow of CAIDA traffic data from 2009-2011 and 2015-2016 were calculated, including number, maximum rate, average rate and average duration of long-lived TCP flows and short-lived TCP flows. It also analyzes the flow characteristic rule and the relationship between various flows, draws the graph line, thus understands the network performance more intuitively.

References

[1]. Thompson K, Miller G J, Wilder R. Wide-area Internet traffic patterns and characteristics[J]. IEEE Network, 2002, 11(6):10-23.

[2]. Mccreary S, Claffy K. Trends in wide area ip traffic patterns: a view from ames internet exchange[J]. In Proceedings of] 3th ITC Specialist Seminar on Measurement and Modeling of IP Traffic, 2000.

[3]. Lee D, Carpenter B E, Brownlee N. Media Streaming Observations: Trends in UDP to TCP Ratio[J]. International Journal on Advances in Systems & Measurements, 2011, 3(3 and 4):147-162.

[4]. Iuoras N, Le-Ngoc T, Ashour M, An IP-based satellite Communication system architecture for interactive multimedia services[C]. International Journal of Satellite Communications and Networking, 2003, 21:401–426.

[5]. T. Benson, A. Akella, D. Maltz, “Network Traffic Characteristics of Data Centers in the Wild” Internet Measurement Conference 261-280, 2010.

[6]. WenJia Fang and Larry Peterson. Inter-AS traffic patterns and their implications. In Proceeding of IEEE GLOBECOM 99, pages 1859{1868, Rio de Janeiro, Brazil, 1999.

[7]. Nevil Brownlee and kc claffy. Understanding internet traffic streams: Dragonflies and tortoises. IEEE Communications Magazine, 2002.

[8]. Shriram Sarvotham, Rudolf Riedi, and Richard Baraniuk. Connection-level analysis and modeling of network traffic. In Proceeding of ACM SIGCOMM Internet Measurement Workshop 2001, pages 99-104, San Francisco Bay Area, November 2001.