Advances in Computer Science Research, volume 89
2019 International Conference on Wireless Communication, Network and Multimedia Engineering (WCNME 2019)

Research on TCP Fairness in Data Center Networks

Bo Zhang[1,*], Tianhang Yin[2], Zhong Wang[3], Junjie Geng[4] and Lei Chen[5]
[1,2,3,4]No. 1 Dingfuzhuang East Street Chaoyang District, Beijing, China
[5]No.10, Huixin Dongjie, Chaoyang District, Beijing, China
[*]Corresponding author

***Abstract*—"Big data" has become a hot spot of the times, and the processing of big data is inseparable from Data Center Networks (DCN). As the infrastructure of the information era, data center networks provide a variety of network services and become key support technologies for future Internet/Cloud computing services and applications. In the process of soliciting data and using data, the performance of the data center network is very important. In the data center network communication process will produce a large number of TCP flows, if you can speed up the transmission of TCP flows, you can greatly reduce the response time, bring economic benefits; the other hand, if you can't properly control the data center network appears With a large number of TCP streams, there will be network congestion. This paper analyzes the problem of TCP fairness in data center network theoretically, designs the network topology according to the characteristics of the data center network transmission process, uses NS2 simulation platform to carry out experiments, and simulates and analyzes TCP fairness according to the principle of maximum and minimum fairness. The relationship between sex and bottleneck bandwidth, network parameters such as RTT, and the number of flows and other factors, and finally gives relevant conclusions.**

Keywords—DCN; TCP; NS2; fairness

I. INTRODUCTION

The data center network is a low-latency, high-bandwidth network that provides users with services such as high-performance computing and mass storage. Low latency means that the RTT time of the data center's internal network is very short, usually microseconds; high bandwidth means that the communication between servers in the cluster has high bandwidth. Currently, the research data center network is of great significance to the bourgeon and application of the cloud computing industry.

"Big data" is a hot spot in the current era, and the processing of big data is inseparable from Data Center Networks (DCN). The data center network is the base installation of the Internet. It is a collection of computing resources, storage resources, and network resources. It provides a wide range of cloud services, such as web search, video on demand, social networking, recommendation systems, cloud storage, and science. Compute, etc. Providing support for these services is a wide variety of distributed big data processing tools that run inside the data center. In order to speed up data processing, these tools often use a partitioning/aggregation mode of work, which divides the task into a large number of sub-tasks, distributes them to multiple servers for parallel computing, and then aggregates the periodic results of each sub-task into the final result [1].

In the process of improving the performance of data center networks, the design of data center networks is constantly making efforts at all levels. There are two major issues that need to be considered in the improvement process. On the one hand, it is necessary to increase the transmission efficiency as much as possible, which is easier to implement by some traditional methods, such as increasing the network bandwidth and optimizing the control algorithms; on the other hand, it is necessary to worry about not being able to "disregard each other": in a data center network, all users should be equal. Bandwidth should be allocated on a requirement. It should not always be queued at the end of the queue because it requires less bandwidth. Even requirements are immediately ignored. It is so frightening. We should try our best to avoid it. However, in fact, the second aspect happens to be ignored by the traditional TCP protocol. In the transmission, the allocation of bandwidth is prone to "more and less", that is, the faster the transport stream will be divided into more bandwidth, and the slower the bandwidth, the less the traffic distribution will be. This creates a vicious circle [2] [3].

At the physical layer, if you want to increase bandwidth, you can use a new topology, use multiple low-end servers, and use the partition/aggregation mode instead of an expensive high-end switch to reduce costs and improve efficiency. The conception at the network level is to make full use of the bandwidth. The main conception is also the partition/aggregation model. It is not same to the physical layer. It divides the TCP flow, divides a TCP flow into multiple sub-flows, and allocates them dynamically, making full use of idleness. path of. Both the physical layer and the network layer can accelerate the transmission of TCP flows. To solve the impartiality problem, it requires to be disposed of the transport layer. The feasible solution is to optimize the TCP protocol according to the characteristics and needs of the data center network.

Therefore, finding out the transmission rule of TCP flow under the many-to-one condition in the data center network is very important for improving the performance of the data center network, and can also provide a clear direction for the subsequent improvement work.

II. THEORETICAL MODEL

We will use the Fairness Index to quantify the fairness of TCP flows in data center networks. The fairness index quantization algorithm is the most commonly used algorithm to measure fairness at the current stage.

According to the maximum and minimum fairness criteria, a fair distribution system must have at least the following characteristics:

(1) All TCP streams that are to use the same link need to be sorted according to their own traffic size, from small to large, and used as the standard for allocating bandwidth.

(2) The bandwidth allocated to each TCP stream must not exceed its required bandwidth;

(3) If the amount of resources required by some flows exceeds the load capacity of the links themselves and cannot meet their requirements, then equaling bandwidth should be allocated to these flows.

However, for traffic control and congestion avoidance, the TCP protocol will continuously increase the congestion window during the transmission process and occupy as much bandwidth as possible until the link is full. Therefore, in the many-to-one mode of the data center network, multiple TCP flows share a single link. Only when a certain rule is used to control the allocation of bandwidth, multiple small flows cannot be guaranteed to affect each other. Some resources, otherwise it is bound to have a certain stream throughput increase, causing the decline in the flow of other streams, resulting in unfair phenomenon [4] [5].

When using the maximum and minimum fairness criteria, we will elicit a fairness index F as the criterion for judgment. Suppose there are n TCP small flows sharing a link X. Assume $\mathbf{x_i}(t)$ represents the portion of bandwidth occupied by stream i at time t. The maximum and minimum fairness index of link X at time t is defined as:

$$F(X,t)=\frac{\left[\sum_{i=1}^{n} x_i(t)\right]^2}{n\sum_{i=1}^{n} x_i(t)^2} \quad (1)$$

The fairness index F is greater than 0 less than 1, and the fairness of bandwidth allocation is obviously affected by F. From the simulation results, with the continuous increase of F, the bandwidth allocation is more equitable. When each TCP stream can be divided into equal bandwidth, the fairness is the highest, F is 1 at this time. On the other hand, if a small stream occupies the entire bandwidth, the rest of the allocated bandwidth is 0, which is the most unfair situation. Under this condition, the final result of F is 1/N.

Even in the TCP design process, high efficiency and high fairness have always been the goal and direction of efforts, but in fact the TCP transmission protocol has not been implemented. Systems with convergence centers have stability, but TCP does not have convergence. Therefore, without changing the existing TCP congestion control strategy, the TCP protocol still has room for optimization. The space for optimization can be divided into two parts: (1) Efficiency. To improve efficiency, an effective way is to reduce the occurrence of packet loss. An indirect way: first let go of some packets, and then discard some packets randomly, so that the receiving end of TCP will have out-of-order packets. The out-of-order packets send a redundant ACK. The sender can process the packet loss through fast retransmission/recovery without having to reduce the congestion window. The sawtooth is refined, unlike the time-out retransmission, the rate drops to low and low all at once. The flow control of the BIC mode can refer to it and refine the window growth mode. In essence, the binary search method is used to gradually approach stability. The rate value then makes a small range of sawtooth oscillations around this stable value instead of using a fixed threshold at the time of congestion avoidance like traditional congestion control to control the multiplicative reduction. (2) Fairness. For fairness, it is to provide RTT type, improve sampling accuracy, and add weighting parameters to adjust the congestion window at the sending end [6].

III. Experimental Verification

A. Experimental Design and Experimental Results

Figure I shows the structure of the simulation topology. We first proceed to the simulation environment with only two TCP flows. Among them, nodes 0 and 1 are data sources, nodes 2 and 3 are bottleneck links, and node 4 is a receiver. The specific parameters are shown in Table I and II. The simulation duration is 125s. The FTP stream is sent from the 0.5s to the 125s. The experimental results are shown in Figure II and III.

TABLE I. LINK PARAMETERS (SIMULATION 1)

Link label	Link bandwidth	Link delay
0-2	2Mb	5ms
1-2	2Mb	5ms
2-3	2Mb	5ms
3-2	2Mb	5ms
3-4	2Mb	5ms

TABLE II. DATA SOURCE PARAMETERS (SIMULATION 1)

Ftp flow label	Window size	Package size
0	8000	512
1	8000	512

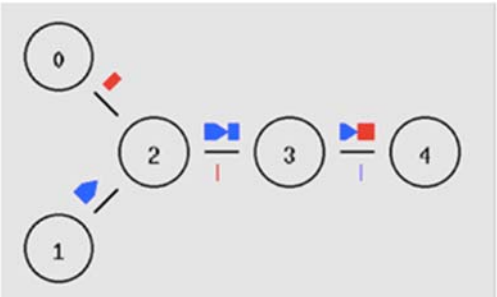

FIGURE I. TWO-SOURCE TOPOLOGY

As shown in Figure II, when the trend becomes stable, the source 0 has a throughput of 311.94, and the source 1 has a throughput of 1691.48. According to calculation formula (1), we can conclude that when the trend is stable:

$$F=\frac{(311.94+1691.48)^2}{2*(1691.48^2+311.94^2)}=\mathbf{0.679} \quad (2)$$

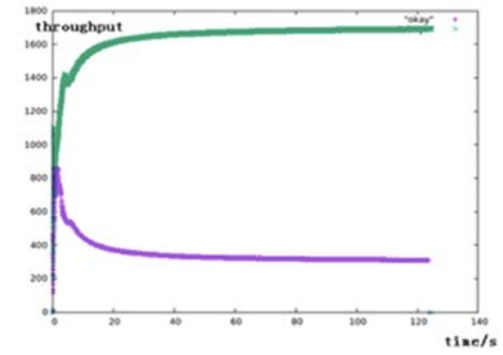

FIGURE II. THROUGHPUT SCHEMATIC (SIMULATION 1)

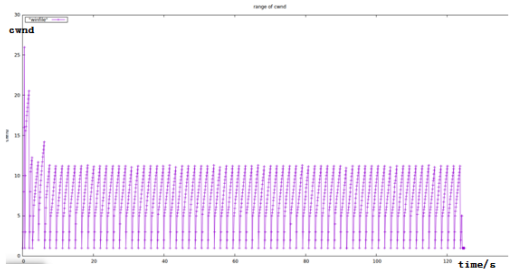

FIGURE III. CWND SCHEMATIC (SIMULATION 1)

After adjusting the parameters, the simulation results are shown in Table III and Ⅳ, repeat experiments and get the results shown in Figures Ⅳ and Ⅴ. As shown in Figure Ⅳ, when it is stable, the throughput of source 0 is 350.33, and the throughput of source 1 is 1650.81. According to the formula (1), we can draw F = 0.70 when it is stable.

TABLE III. LINK PARAMETERS (SIMULATION 2)

Link label	Link bandwidth	Link delay
0-2	2Mb	5ms
1-2	2Mb	5ms
2-3	2Mb	10ms
3-2	2Mb	10ms
3-4	2Mb	5ms

TABLE IV. DATA SOURCE PARAMETERS (SIMULATION 2)

Ftp flow label	Window size	Package size
0	8000	512
1	8000	512

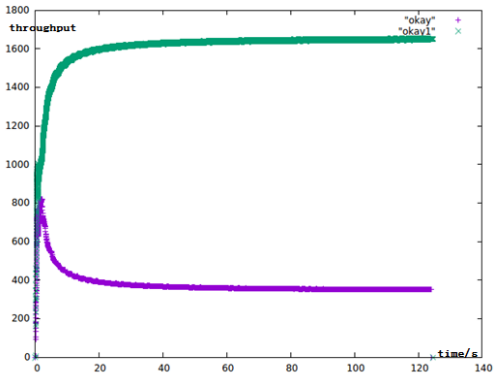

FIGURE IV. THROUGHPUT SCHEMATIC (SIMULATION 2)

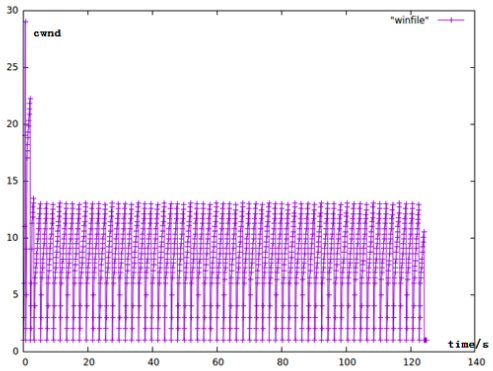

FIGURE V. CWND CHANGES (SIMULATION 2)

Similarly, as shown in Figure Ⅵ, we can simulate the conditions of the three sources. Adjust the parameters as shown in Table Ⅴ and Ⅵ, and repeat the experiment. The result is shown in Figure Ⅶ.

By calculation, it can be concluded that F = 0.76 for three sources.

TABLE V. LINK PARAMETERS (SIMULATION 3)

Link label	Link bandwidth	Link delay
0-2	2Mb	5ms
1-2	2Mb	5ms
2-3	2Mb	5ms
3-2	2Mb	5ms
3-4	2Mb	5ms
5-2	2Mb	5ms

TABLE VI. DATA SOURCE PARAMETERS (SIMULATION 3)

Ftp flow label	Window size	Package size
0	8000	512
1	8000	512
5	8000	512

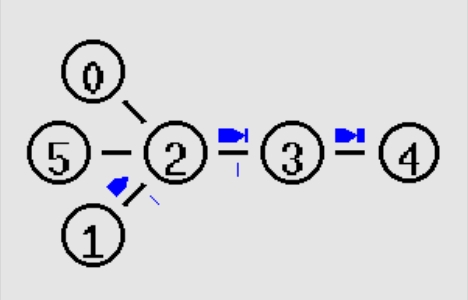

FIGURE VI. THREE-SOURCE TOPOLOGY

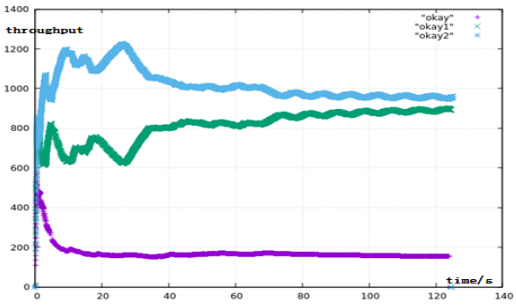

FIGURE VII. THREE-SOURCE THROUGHPUT

B. Analysis of Experimental Results

From a horizontal perspective, in the course of each data transmission, the fairness of TCP evolves from fairness to more and more unfairness after a certain amount of oscillating time.

In the longitudinal direction, the bandwidth and delay of the bottleneck link also affect the fairness. In simulation experiments, the longer the delay, the wider the bandwidth and the better the fairness of TCP. At the same time, according to the change of CWND, it can be seen that the delay and bandwidth of the link also affect the fairness oscillation time. Under the same experimental conditions, the more sources, the better the fairness of TCP.

In fact, due to the characteristics at the beginning, the TCP itself is inefficient and fair. The inequity of TCP is also caused by the TCP congestion control mechanism. It is precisely because all the streams after initialization share a bandwidth that leads to this result. TCP adjusts bandwidth allocation according to the current transmission performance. The CWND window also increases until the packet loss occurs. For each source, it will tentatively increase its own transmission rate, increase rapidly, and it will suddenly decrease after the packet loss occurs, and repeat the previous process. However, when one of the roads has an advantage, because of TCP's congestion control mechanism, it needs a "preferential public," and the slowing down rate is a process that is likely to benefit others. The TCP protocol encourages this kind of thing to happen, but it is in the process of adjustment, other paths cannot be compared with it, but will only be squeezed more and more, at a slower and slower rate, and the throughput will be getting smaller and smaller. Therefore, there will be a situation where one road is increasingly crowded for other roads and the fairness is getting worse.

IV. SUMMARY

In this paper, in the course of each data transmission, the fairness of TCP evolves from fairness to more and more unfairness after a certain oscillating time. The fairness of TCP is also affected by the following two factors: 1. the bandwidth of the bottleneck link. 2. time delay. A general conclusion can be drawn from the simulation results. The longer the delay and the wider the bandwidth, the better the fairness will be. At the same time, according to the change of CWND, it can be seen that the delay and bandwidth of the link also affect the fairness oscillation time. Under the same experimental conditions, the more sources, the better fairness TCP shows, but at the same time, the transmission rate of some sources is limited to a very low level, although the average is good overall, but this Sacrifice should not be allowed in practical applications.

ACKNOWLEDGMENT

This work was supported by the Fundamental Research Funds for the Central Universities and National College Students' innovation and entrepreneurship training program.

REFERENCES

[1] Chowdhury M, Zaharia M, Ma J, Managing Data Transfers in Computer Clusters with Orchestra. In: Proc of SIGCOMM. New York, NY, USA: ACM, 2011.

[2] Phanishayee A, Krevat E, Vasudevan V, Measurement and analysis of TCPthroughput collapse in cluster-based storage systems, Usenix Conference on Fileand Storage Technologies, 2008.

[3] Zhang J, Ren F, Lin C, Modeling and understanding TCP incast in data centernetworks, INFOCOM, 2011.

[4] https://blog.csdn.net/dog250/article/details/585417722.

[5] Alizadeh M, Atikoglu B, Kabbani A, Data center transportmechanisms: congestion control theory and IEEE standardization, Proceedings of the 46th Annual AllertonConference on Communication, Control, and Computing,Urbana-Champaign, USA, Sep 23-26, 2008.

[6] Yajun Yu, Zheng Liu, Mingwei Xu, Research on TCP Incast in Data Center Network, Computer Science and Exploration, 2017, 11 (09): 1361-1378.