Part3

Junru Xia(jx17) & Kejun Liu(kl50)

9/27/2017

Introduction

In this document, we will explain how we measure the bandwidth independent delay across two CLEAR servers using our ping-pong client/server software, and estimate the bandwidth of the network link.

Generally speaking, the data we get from our ping-ping client/server software is the total latency time(T), which is the sum of the data size and bandwidth dependent transmission delay(TD) and the data size and bandwidth independent delay(ID). Besides, the data size and bandwidth dependent transmission delay approximately equals to data size (S) divided by bandwidth(W). We can use these formulas to represents the relationship.

$$T = TD + ID$$

$$TD = \frac{S}{W}$$

$$T = \frac{S}{W} + ID$$

Since bandwidth(W) and data size and bandwidth independent delay(ID) are constants, and the data size and total latency time are variables, to simplify the relationship between these variables and constants, we can use this formula.

$$T = aS + ID$$

$$a = \frac{1}{W}$$

Since here we use a linear model, all we need to do is to estimate a and ID. In order to estimate, we should collect data first.

Data Collection

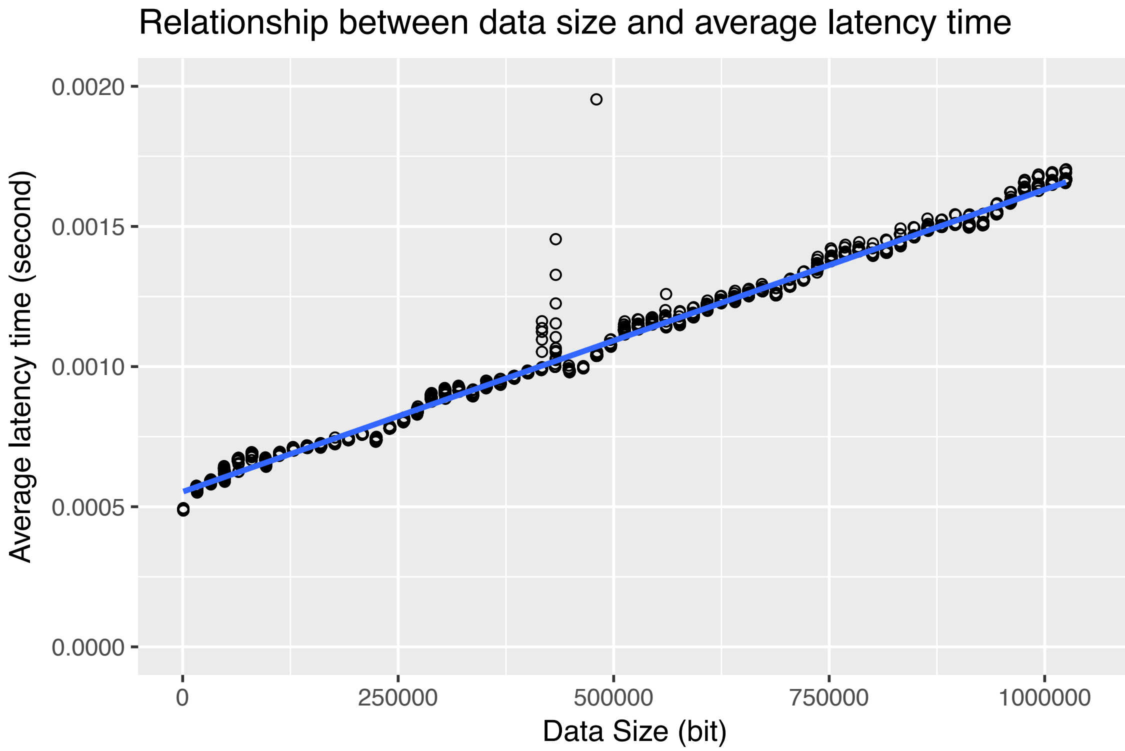
In our ping-pong client/server software, client program should take 4 command line parameters, two of them are data size and count.

In order to increase accuracy, a large quantity of data is needed. Therefore, we wrote a python script to collect data sets automatically. Specifically, in our experiments, data size ranges from 50 to 65535 step by 1000, and count ranges from 200 to 10000 step by 1000. For each run, the total time latency retrieved from client is recorded in out.txt file. The script we used is:

```python
import os
for size in range(50, 65535, 1000):
   for count in range(200, 10000, 1000):
     os.system("./client_num ring.clear.rice.edu 18021 "+str(size)+" "
     +str(count)+" >> ./out.txt")
```

Data Visualization

After collecting our data, we visualize them to see the relationship between data size and time.



From the plot, we can see that most data lies in a certain line, which can prove our hypothesis in the introduction part. So, we can use linear regression to deal with this problem.

Model Establishment

Till now, our model is

$$T = aS + ID$$

$$a = \frac{1}{W}$$

So next, we should use linear regression to estimate a and ID based on data we collected just now. The results are as follows:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.91992e-04	1.134051e-05	43.3836	0
output$size	1.25500e-09	1.800000e-11	70.7993	0

From the output, we can get

$$T = 0.000000001255 * S + 0.000491992$$

Besides, according to the output, we can see that the p value is very low, and R square is quite high, which means we can accept our hypothesis.

Therefore, we can know the **on average**, the bandwidth independent delay is about $0.000491992s$, and the bandwidth should be

$$a = \frac{1}{0.000000001255} = 796812749bps$$

Conclusion

First, in theory, we can see that data size and average latency time have a linear relationship. Besides, visualizing data we collected can prove this hypothesis. So we use linear regression to fit the model, and get the final results.

From our method, we can see that **on average** the bandwidth independent delay is about $0.000491992s$, which is $0.49ms$, and the bandwidth should be $796812749bps$, or $797Mbps$.