Relationship between Network Traffic Topology Map and Degree of Competition

Hanyue He*

International School, Beijing University of Posts and Telecommunications, Beijing 100876, China

* Corresponding author e-mail: 2017213127@bupt.edu.cn

Abstract. This project will examine the relationship between the degree of competition between access providers and the price of Internet access in a particular country, as well as the impact of access provider topological characteristics on price and quality. Its purpose is to understand whether greater competition will result in lower network access prices. This project uses a scatter plot to show the relationship between these variables and attempts to build a model to express this relationship. The data used in this project includes Internet access prices and topology information of Internet Service Providers (ISPs). The data set is from the author's previously published article 'Establishing a database of competition between ISPs based on Python language'. The purpose of this paper is to demonstrate the relationship between the price and the bandwidth of the Internet access and the competition among access provider by the method of data analysis. And at the end, we increase the accuracy of the model by adding the possible influence variables.

1. Introduction

As we enter the era of web 2.0, the role of the Internet in our lives is becoming more and more important. This paper discusses the relationship between network access prices and the degree of competition of ISPs. The main point of the paper is to use methods from data analytics to study the relationship between network competition and Internet access prices or bandwidth. Try to find as many variables as possible that could affect the price or bandwidth. For instance, these variables include the degree of competition among ISPs, topology information of ISPs, and the number of customers of each ISP. Based on the relationship between these variables and the network access price, we can make reasonable assumptions. For example, the more intense the competition, the lower the network access price. The greater the number of Peering interconnections, the lower the network access price. The greater the number of Transit interconnections, the higher the network access price and so on. What we want to do is that use data and models to scientifically validate our assumptions and hypothesis. The datasets used in this paper were created by the author in previous papers and research.

2. Background

2.1. Topology information

Topology information of ISPs could have some influences on Internet access prices as well. Here, topology information means different interconnection between ISPs. One of them is peering, the other is transit interconnection. Peering interconnection has fewer costs in comparison to transit interconnections. The reason why peering interconnections have fewer costs will be discussed in section

3.1. In general, if a user wants to interconnect with other users all around the world, it is necessary to use some other networks to forward the traffic. For instance, User1-AS1-AS2-AS3-User2, interconnections among those ASes can be peering or transit. That is why topology information has impacts on internet access prices. Here, AS means autonomous system. An AS is a unit that has the authority to decide autonomously which routing protocol should be used in the system. An autonomous system will assign a global unique 16-digit number, and we call this number the Autonomous System Number (ASN) [1].

2.2. Purchasing power parity
In this paper, we need to compare Internet access prices in different countries. We need to use the network access price information calculated by purchasing power parity. Because purchasing power parity has taken into account the living standards of each country. The price of network access using the purchasing power parity is comparable. Purchasing power parity is an equivalent coefficient between currencies calculated according to different price levels of different countries. The purpose is to make a reasonable comparison across countries[2].

3. Evaluation

3.1. Topology information
The topology data information for this project is derived from CAIDA. The website is http://www.caida.org/research/topology/. This site contains a lot of datasets about topology information, such as The IPv6 Topology Dataset, and AS Relationships. The topology information used in this project comes from the AS Relationships dataset.

The business agreement between ISPs is complex, including three common types, provider-to-customer p2c, peer-to-peer (p2p), and sibling-to-sibling (s2s). [3]

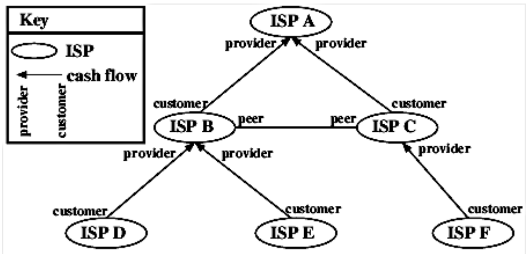


Figure 1. Types of AS relationships [4]

In Figure 1, ASes at lower levels have to pay ISPs at higher levels in order to connect to the rest of the Internet. For example, if D wants to exchange information with C, D has to pay B a fee. B is in this connection plays the role like an intermedia, and D is the customer of B. However, the connection between B and C is a p2p interconnection, which means it will cost less on transit. The premise of p2p interconnection is that the interconnection is equivalent to the interests of both parties, which can save cumbersome traffic records and save costs. The two sides reached a peer-to-peer interconnection agreement on the basis of balanced interests, which is a completely mutually beneficial business behavior.

For ISPs, different interconnections have different costs. In this project we need to count how many p2p interconnections each ISP has and how many p2c interconnections there are. The more p2p interconnections mean the lower the cost, the more p2c interconnections mean higher the cost.

The format of the topology data obtained from CAIDA is: <provider-as>|<customer-as>|-1<peer-as>|<peer-as>|0. For example, 1|3|-1 means that AS3 is a customer of AS1. We can find that the format represented by x|x|-1 is p2c, and the format represented by x|x|0 is p2p. Therefore, we only need to count the number of p2p and p2c interconnections of an AS. However, there is only AS NUMBER included in this dataset such as AS3. In order to combine with the dataset that we have processed before, we need to get the ISP name corresponding to the AS NUMBER. From the website http://as-rank.caida.org/asns, we can get them.

When processing this data set, each row of data is segmented with '|' as a separator. Traverse the data of each row, set the appropriate conditional statement, and write the processed data to the file. Similarly, the data is statistically calculated using DataFrame, and the number of p2p and p2c interconnections for each AS is calculated.

3.2. Plots and models

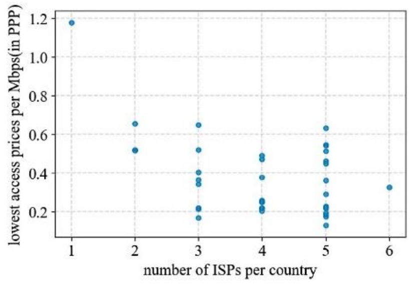
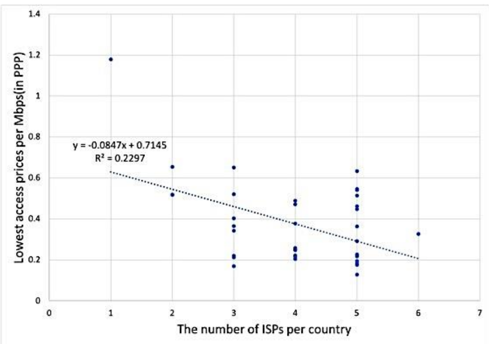


Figure 2. Scatter Plots about lowest access prices and number of ISPs per country

Figure 2 is a scatter plot about the minimum Internet access price and the number of ISPs in each country. Each point in the figure represents a country. The vertical axis of Figure 2 represents the lowest Internet access price per country. The horizontal axis of Figure 2 represents the number of ISPs in each country. As mentioned above, the number of ISPs in each country can represent the degree of competition in each country. As can be seen from the figure, the overall trend is that the higher the degree of competition, the lower the minimum Internet access price. Liechtenstein has only 1 ISP (Telecom Liechtenstein), while the lowest access price of it is the highest among 38 countries (1.1788569853086). Denmark has the largest number of ISPs (6 ISPs). Although Liechtenstein's access prices are much higher than in Denmark, ISPs in Denmark offer higher bandwidth than Liechtenstein. This shows that the more intense competition can lead to lower prices and higher speeds to some extent.

(left)


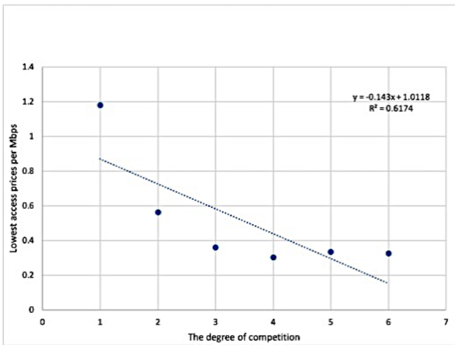
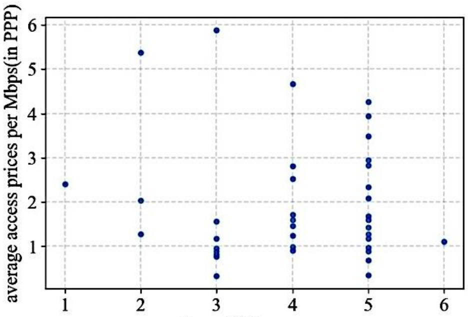
(right)


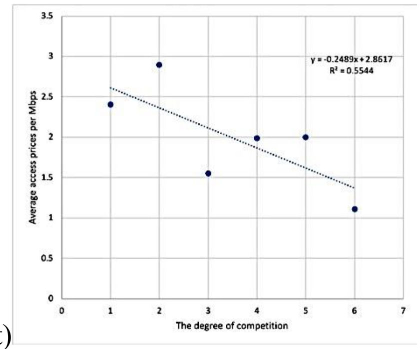
Figure 3. Linear regression model about lowest access prices and competition

According to the trend presented in Figure 2, a linear regression model was established, and the value of R^2 was calculated.

The accuracy of the regression model shown in Figure 3 is not high enough. After analysis, this is because many countries have the same number of ISPs, and the Internet access prices provided by these ISPs are scattered. For example, when ISPs=5, there are 16 scatters distributed on the vertical axis. However, when ISP=6, there is only one scatter. This has a great impact on the accuracy of the model. Therefore, we can divide the data into 6 groups according to the number of ISPs, which are 1-6. Calculate the average value of Internet access



(left)



(right)

Figure 4 Scatter Plots about average access prices and number of ISPs per country

prices in each group. After verification and re-establishment of the model, the accuracy of the model has indeed improved. Figure 3.1 shows the newly established linear regression model. The value of R^2 improves from 0.2297 to 0.6174.

In Cyprus, there are 3 ISPs. The average access prices per Mbps of it is the highest among 38 countries. This is because the bandwidth those ISPs provided is lowest. In Romania, the average bandwidth is the highest among 38 countries. From Figure 4, a clear trend cannot be seen. Therefore, similarly, we have done the same treatment for the average access prices like processing lowest access prices data, and the model is built as shown below:

From Figure 3.1 and 5.2, it is easy to find that there is a similar trend. Intense competition will result in the decline of access prices.

3.3. Improvement of model accuracy

To improve the accuracy of the model, the project used more years of data to optimize the model. The amount of data nearly doubled. In above figures, there is no significant relationship between bandwidth and competition. However, by adding data from 2017, it produced a different result. The results show that the competition degree has an effect on the bandwidth, the more intense the competition degree is, the higher the bandwidth is.

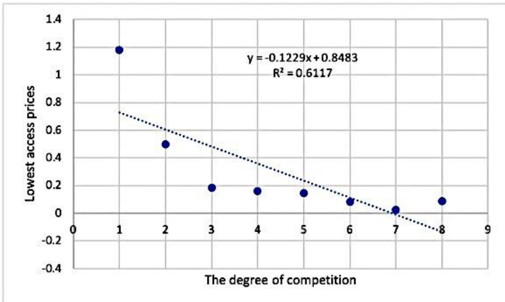


Figure 5 Optimised lowest prices model

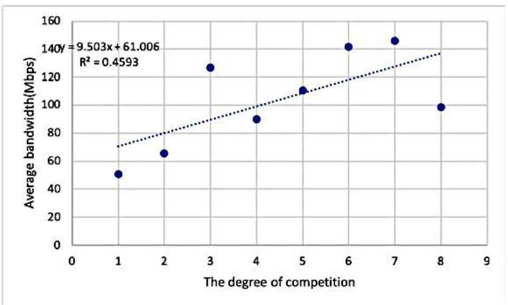


Figure 6 Bandwidth model

Journal of Physics: Conference Series **1449** (2020) 012082 doi:10.1088/1742-6596/1449/1/012082

4. Conclusion

In conclusion, it can be seen from established models that the degree of competition among ISPs has a huge impact on Internet access prices. The higher degree of competition can result in lower Internet access prices and better network quality.

What's more, topology information also has influence on the prices. More p2p interconnections an ISP have, lower the Internet access prices. On the contrary, more p2c interconnections an ISP have, higher the access prices. With more and more years of data digging, the accuracy of the model is improved.

References

[1] En.wikipedia.org. (2019). Autonomous system. [online] Available at: https://en.wikipedia.org/wiki/Autonomous_system.

[2] En.wikipedia.org. (2019). Purchasing power parity. [online] Available at: https://en.wikipedia.org/wiki/Purchasing_power_parity.

[3] Analysis, C. (2019). AS Relationships. [online] CAIDA. Available at: http://www.caida.org/data/as-relationships.

[4] Analysis, C. (2019). AS Relationships. [online] CAIDA. Available at: http://www.caida.org/data/as-relationships