Planning Methodology for Mapping between Network Bandwidth Capacity and Physical Traffic

Wonhyuk Lee, Minki Noh, Dongkyun Kim, Byoungyeon Park and Seunghae Kim

Korea Institute of Science and Technology Information
{livezone, mknoh, mirr, bypark, shkim}@kisti.re.kr

Abstract

Optimal resources allocation should be done in first establishment stage. But, optimal capacity allocation in consideration of practical standards such as current traffic volume has a significant impact on performance. In this paper, we've attempted to model an appropriate mapping method based on network traffic as the basis for early allocation policy.

Keywords: *Bandwidth Capacity, Traffic Mapping*

1. Introduction

It is very important to secure an optimum capacity of network backbone and provide services in terms of economic aspect and work efficiency. In particular, virtual organization has even been used in forming a research network. In fact, it is hard to find a paper on the estimation of network capacity even though it is related with logical groups and application groups such as high-energy physics and high-performance computing resource linkage. In the past, the total capacity of network backbone was allocated to a user network on a static manner. At present, it is a trend to study technology which is necessary to realize environment in which bandwidth or resources could be dynamically allocated by a user such as Software Defined Network (SDN). However, it is still difficult to apply perfect dynamic allocation technology for establishment of backbone infrastructure and allocation of resources. In particular, optimal resources allocation should be done in an early stage. Therefore, optimal capacity allocation in consideration of practical standards such as current traffic volume has a significant impact on performance. This paper has attempted to model an appropriate mapping method based on network traffic as the basis for early allocation policy.

2. The Background of Estimation-based Network Capacity Model

2.1. Capacity Estimation Method

In terms of a way to estimate network capacity, there is a questionnaire survey. This method is generally used by the Internet service providers to estimate the number of service subscribers and subscription rates based on the developed product models. The factors which determine network traffic include the number of service subscribers, subscription rates and service distribution rates. After tracking down and analyzing the past records, group meeting, questionnaire survey, opinion poll and market research are conducted.

This paper has attempted to test self-similarity based on the real traffic information and propose an appropriate capacity estimation methodology through the traffic estimation algorithm.

2.2. Traffic Self-similarity

There are various ways to test traffic self-similarity. Even though a current trend could be briefly mentioned, it has a disadvantage of low accuracy. In this paper, the traffic self-similarity was tested using R/S plot. The R/S plot is a common method used to measure the strength of self-similarity by the effect of hurst.

Using the characteristics that R/S of the adjusted scope increases according to the rules of repeated square with H , it is defined as the value of $\log(R(n)/S(n))$ on $\log(n)$. Where, n is a natural number which represents sampling interval in data points while $R(n)$ and refer to the statistical number of process and standard deviation of sample respectively. If it is assumed that the mean of the sample is $\overline{X}(n)$, the equations on W_k and $R(n)/S(n)$ can be defined as follows:

$$W_k = (X_1 + X_2 + \cdots + X_k) - k\overline{X}(n), (k \geq 1)$$
$$R(n)/S(n) = [\max(0, W_1, W_2, \cdots, W_n) - \min(0, W_1, W_2, \cdots, W_n)]/S(n)$$

2.3. Probability Plot (P-P Plot) for Normality

Let's say the order statistics of probability distribution ($X_1, \cdots, X_n$) is $X_{1:n}, \cdots, X_{n:n}$, and the Cumulative Distribution Factor (CDF) of the probability factor X is $\Phi(x)$. Then, let's assume $U_i = \Phi(X_i)$. According to the Probability Integral Transformation, U_i follows $UNIF(0,1)$. After that, let's say the order statistics of $U_1, \cdots, U_n$ is $U_{1:n}, \cdots, U_{n:n}$. Then, $U_{i:n} = \Phi(X_{i:n})$. Here, the Probability Density Function (PDF, g_i) of the order statistics ($U_{i:n}$) can be induced as follows:

$$\begin{aligned} g_i(u_{i:n}) &= \frac{n!}{(i-1)!(n-i)!}\left[\int_{-\infty}^{u_{i:n}} f(x)dx\right]^{i-1} f(u_{i:n})\left[\int_{u_{i:n}}^{\infty} f(x)dx\right]^{n-i} \\ &= \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)}(u_{i:n})^{i-1}(1-u_{i:n})^{n-i} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(u_{i:n})^{\alpha-1}(1-u_{i:n})^{\beta-1} \end{aligned}$$

Where, $f(x)$ is the PDF of $UNIF(0,1)$ distribution, and $\alpha = i, \beta = n-i+1, 0 \leq u_{i:n} \leq 1$. The induced PDF g_i is equal to the PDF of BETA distribution. Therefore, it can be induced that $U_{i:n}$ follows $BETA(i,n-i+1)$. Then, let's say the sampling distribution function is $\Phi_n(\cdot)$. The, $\Phi_n(X_{i:n})$ is $\frac{i}{n}$. After that, the P-P Plot can be defined as graph which expresses

$$(\Phi(X_{i:n}), \Phi(X_{i:n}))$$

on the (x, y) using the said properties. Then, $\Phi(X_{i:n})$ follows $BETA(i,n-i+1)$. Therefore, the expected value is

$$E[\Phi(X_{i:n})] = \frac{i}{n+1}$$
.

Where, the graph will be a line with a slope of $\frac{n+1}{n}$. The deviation from the line is used as the scale for normality test.

3. Network Capacity Estimation Model

The logical network which consists of virtual organization is a virtual network which unites computing resources to solve a common problem or achieve a common goal. It is a very important concept which started from the grid computing for use of multiple system resources. Right now, it is even used in configuration of logical group network to support advanced application research activities using the science technology research network as well as grid computing. For this, it is an essential process to properly allocate logical links and adjust the allocated logical network because bandwidth is a limited resource. In other words, it is required to estimate network capacity in consideration of VO.

3.1. Method of Capacity Estimation

In the physical network in which multiple logical links are allocated to configure VO, the target of capacity estimation is a single physical link which includes multiple logical links. The purpose of capacity estimation is to diagnose traffic of physical link and count the number of the logical links which can be additionally allocated to the physical link. It can change depending on the consumption and distribution of the already allocated logical links (or the logical links to be allocated soon). In other words, the results of the capacity estimation are not constants. Because they can vary depending on the variables of network circumstances, it is necessary to measure and simulate traffic by logical link for capacity estimation. If traffic is simulated, distribution and mean could be critical factors. In addition, measured or created data should be sufficient enough to have representative.

3.2. Capacity Estimation Algorithm

The goal of this algorithm is to reduce a scope of investigation through alignment, section division and index adjustment. This algorithm requires the following information:

① No. of logical links
② Bandwidth of physical links
③ No. of traffic data by VO

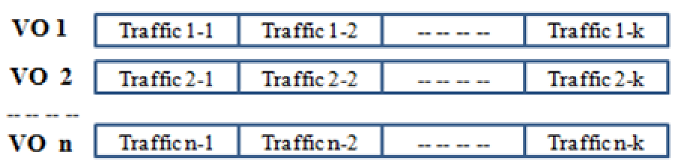


Figure 1. Structure of Classified Traffic Data

④ Traffic data

- Stage I: Alignment

In Figure 11 as shown above, ascending sort by VO is carried out. Even though it is a simple arrangement, it can enhance overall algorithm efficiency if Quick Sort or Merge Sort is used.

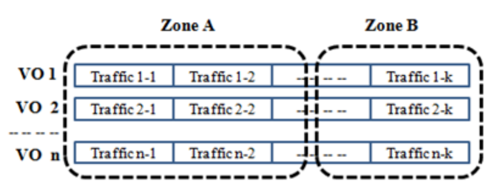


Figure 2. Zone Dividing Model

- Stage II : Domain Division

To minimize and simplify search process, data set was divided into two zones. 'Zone B' exceeds bandwidth even though the sum is calculated by combining traffic in each VO in any manners. On the contrary, 'Zone A' does not exceed bandwidth even though the sum is calculated by combining traffic in each VO in any manners. In other words, if selected within 'Zone A,' it is not possible to make a combination which exceeds bandwidth. In the following two cases, traffic combination in each VO can exceed bandwidth.

① Combination in Zone B
② Combination in which both zones are used

The zones are divided by searching for the spot where Zone B begins. It proceeds from the left to the right. The row whose sum first exceeds bandwidth is found.

- Stage III : Number of Cases

In two cases in which the traffic combination in Stage II can exceed bandwidth, the number of cases is estimated.

- Stage IV: Combination in Zone B

If the number of data by VO in Zone B is 'x,' the number of possible combinations is as follows:

$${}_xC_1 \times {}_xC_1 \times {}_xC_1 \times \cdots \times {}_xC_1 = x^n$$

- Stage V: Combination in which both zones are used

To perform the process, the indexes by VO are necessary. The following process is repeated through the index control:

Table 1. Process of Adjusting Index

① Each index advances one column by one column by creating precise combinations without missing one starting from the first data on the left toward Zone B.
② Check if all indexes are located in Zone B. If they do, the following processes should not be repeated because they are the same with the Stage IV.
③ If the indexes indicate both zones, check if the sum of the data indicated by the indexes exceeds bandwidth.

- Stage VI: Estimation of final probability

As shown in the Clause 2 Application of Capacity Estimation, probability can be obtained by dividing the number of combinations which exceed bandwidth by the number of all possible combinations in logical links. The number of combinations which exceed bandwidth is the sum of Stage IV and Stage V. If the number of cases after performing the Stage III-2 is 'y,' the probability is as follows:

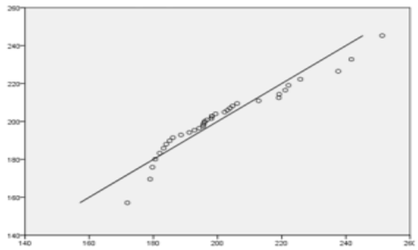
$$\therefore \text{Probability} = \frac{x^n + y}{k^n} \quad (0 \le x \le k, 0 \le y \le k^n)$$

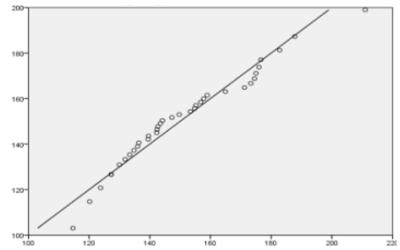
4. Simulation

4.1. Traffic Measurement & Analysis

At present, various organizations are linked to the joint-use-of-super-computing network of KREONET. Then, traffic similarity by organization was analyzed.

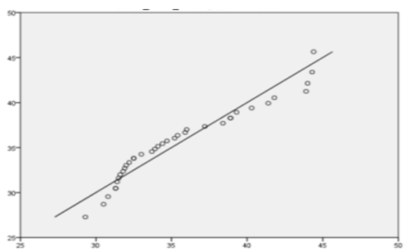
4.1.1. Institute A (Q-Q Plot of Institute A (In/Output))

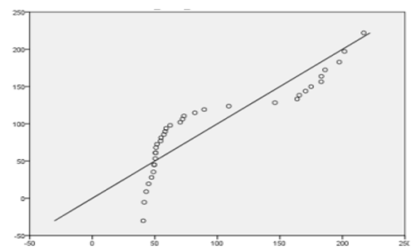




		Input	Output
		Laplace Distribution	Logistic Distribution
Optimal Probability Distribution	Location	201.186	151.028
Parameter of Probability Density Function	Modulation Analysis	13.094	11.860

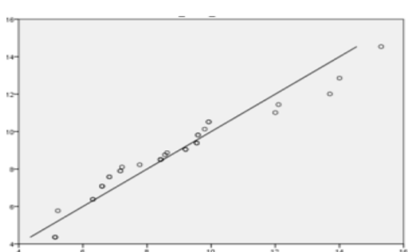
4.1.2. Institute B (Q-Q Plot of Institute B (In/Output))

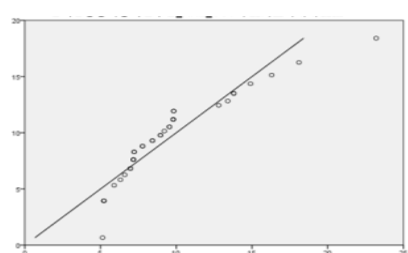




		Input	Output
		Reg. Log Distribution	Reg. Distribution
Optimal Probability Distribution	Location	35.287	109.896
Parameter of Probability Density Function	Modulation Analysis	0.122	1.814

4.1.3. Institute C (Q-Q Plot of Institute C (In/Output))





		Input	Output
		Laplace Distribution	Logistic Distribution
Optimal Probability Distribution	Location	8.6081	9.540
Parameter of Probability Density Function	Modulation Analysis	1.762	2.195

4.2. Traffic Estimation with Random Number

The fabricated algorithm cannot get results without a certain amount of data because traffic is estimated with a statistical technique instead of through the fixed number-based traffic estimation. In other words, the status can be diagnosed by capturing the existing traffic, or the network can be estimated by estimating and randomly generating traffic. In this section, the random numbers which followed Normal Distribution, Lognormal Distribution and Poisson Distribution were created, and the results were analyzed in consideration of diverse circumstances. Microsoft Excel 2010's random generation functions were used, and the circuits with three logic networks were simulated.

4.2.1. The Case when Bandwidth is Sufficient

Table 2. Sufficient Bandwidth with three Logical Networks

Num. of VO	VO1	Normal	m	σ	Probability
3			370	5	
Num. of data	VO2	Lognormal	m	σ	0.90%
100			5	0.4	
Bandwidth	VO3	Poisson	λ		
1 Gb			200		

There are three logical links which are comprised of VO, and 100 traffics were randomly created by link. The bandwidth in physical link is 1Gbps. With the traffic data of VO1, random numbers under the normal distribution (mean: 370, standard deviation: 5) were created. Using the traffic data of VO2, random numbers under the normal distribution (mean: 5, standard deviation: 0.4) were created. Then, they were under EXP() operation and lognormal distribution. With the traffic data of VO3, in addition, random numbers under the Poisson Distribution (mean: 200) were created. Here, the probability that the sum of traffic in logical links exceeds the bandwidth of physical links is 0.90%. In other words, the physical links could be overloaded for up to 13 minutes a day if things get worse.

4.2.2. The Case when Bandwidth is Insufficient

Table 3. Insufficient Bandwidth with Three Logical Networks

Num. of VO	VO1	Normal	m	σ	Probability
3			380	5	
Num. of data	VO2	Lognormal	m	σ	8.22%
100			5	0.6	
Bandwidth	VO3	Poisson	λ		
1 Gb			300		

There are three logical links which are comprised of VO, and 100 traffics were randomly created by link. The bandwidth in physical link is 1Gb. With the traffic data of VO1, random numbers under the normal distribution (mean: 380, standard deviation: 5) were created. Using the traffic data of VO2, random numbers under the normal distribution (mean: 5, standard deviation: 0.6) were created. Then, they were under EXP() operation and lognormal distribution. With the traffic data of VO3, in addition, random numbers under the Poisson Distribution (mean: 300) were created. Here, the probability that the sum of traffic in logical links exceeds the bandwidth of physical links is 8.22%. In other words, the physical links could be overloaded for up to 1 hour and 58 minutes a day if things get worse.

5. Conclusion

This paper has investigated how to analyze traffic in large-capacity research network and application system statistically. If the results of the measurement and analysis on traffic are compared, and similarities between two traffics are tested after creating simulated traffic and applying it to simulation, it will be possible to establish a technical ground to measure similarities between real and simulated traffics.

In fact, it is very difficult to measure network performance by creating large traffic. Therefore, network performance is measured using a simulation traffic generator. The traffic similarity test algorithm can enhance the performance of the simulation traffic generator.

It is likely that the results of this paper would be used in estimating capacity on mobile traffic and mobile network. It appears that this paper would be able to suggest a way to use the resources which are used in various R&D activities more efficiently using statistics.

References

[1] C. Dovrolis, P. Ramanathan and D. Moore, "Packet Dispersion Techniques and a Capacity Estimation Methodology", IEEE/ACM Transactions on Networking, **(2004)**.

[2] M. Jain and C. Dovrolis, "Ten fallacies and pitfalls on endto-end available bandwidth estimation", in: Proceedings of ACM Internet Measurements Conference, Sicily, Italy, **(2004)**.

[3] T. Tunali, A. Kantarci and N. Ozbek, "Robust Quality Adaptation for Internet Video Streaming, ultimedia Tools and Applications", vol. 27, Springer, Berlin, **(2005),** pp. 341–357.

[4] J. Sommers, P. Barford and W. Willinger, "SPLAT: A visualization tool for mining Internet easurements", in: Proceedings of Passive and Active Measurement Conference, **(2006)**.

[5] T. Tsugawa, C. L. T. Man, G. Hasegawa and M. Murata, "Inline bandwidth measurements: implementation difficulties and their solutions", in: Proceedings of the Fifth IEEE/IFIP Workshop on End-to-End Monitoring Techniques and Services, **(2007)**.

[6] J. Prokkola, M. Hanski, M. Jurvansuu and M. Immonen, "Measuring WCDMA and HSDPA delay characteristics with QoSMeT", in: Proceedings of the IEEE International Conference on ommunications (ICC), **(2007)**.