Quantitative Analysis of the Correlation of Round-Trip Times between Network Nodes

Nafei Zhu[1,2], Jingsha He[1*], Yue Zhou[2] and Wei Wang[2]

[1]*Beijing University of Technology, Beijing 100124, China*
[2]*National Application Software Testing Labs, Beijing 100193, China*
zyy@emails.bjut.edu.cn, jhe@bjut.edu.cn

Abstract

Round-trip time (RTT) is defined as the round-trip delay of a data packet between a source and a destination node and is a very important metric for measuring the performance of network applications. The measurement and estimation of RTTs has received a great deal of attention in network measurement research over the years. An interesting question is how RTTs from different source nodes to a common destination node are correlated, which describes to what extent the RTTs exhibit the same changing trend. Such correlation is useful since it would make it feasible to only make some nodes perform RTT measurement while some others estimating RTTs based on the correlation, thus reducing the overhead incurred by RTT measurement. In this paper, we perform a quantitative analysis on the correlation of RTTs, in particular, how network topological properties can affect such correlation. Based on experiment data, our analysis shows that (1) correlations are not much different for path length combinations that only differ in the length of the shorter private path but will become the weakest in statistics when the lengths of the two private paths become the same; (2) correlations are not much different for path length combinations that have the same path length ratio (which is defined as the ratio of the length of the common path to that of the longer private path); and (3) correlation increases along with path length ratio with ratio 1/1 being the inflection point. The results from our analysis of RTT correlation can be applied not only to RTT measurement and estimation, but also to the deployment or selection of measurement nodes in a general network measurement infrastructure.

Keywords: *Internet; Measurement; Round-trip time; Correlation*

1. Introduction

Round-trip time (RTT) is a very important network metric to measure the performance and quality of network applications such as VoIP and has thus become a basic parameter for network measurement. Consequently, understanding the correlation of RTTs between network nodes is very important for network measurement since such correlation can be used to allow some nodes to estimate RTTs based on the results of RTT measurement by some other nodes, thus reducing the overhead incurred by network measurement. It can also be used to determine where to best place measurement nodes to perform the measurement. Although RTTs between network nodes can be affected by topological factors such as path length and runtime factors such as path load, and thus the correlation, in this paper, we mainly focus on analyzing

* The corresponding author (email: jhe@bjut.edu.cn).

the effect of path length on such correlation under several representative levels of network load. Our analysis relies on the experiment data generated using the simulation tool OPNET [1].

There has been some previous work on the study of network distance prediction which is related to our work. Francis, *et al.*, proposed an architecture called IDMaps for estimating the latency of arbitrary network paths from some known nodes by running a spanning tree algorithm over the received distance map [2]. Chen, *et al.*, proposed a system called Internet Iso-bar to cluster hosts based on the correlation of their perceived network distance for an overlay distance monitor system [3]. Agarwal and Lorch designed a latency prediction system for game match-making scenarios with one novel feature being the synthesis of geo-location with a network coordinate system [4]. Chen, *et al.*, designed a system called Pharos that leverages multiple coordinate sets at different distance scales so that one scale will be chosen each time for the prediction in order to achieve a high level of accuracy [5]. Liao, *et al.*, formulated the problem of network distance prediction as a problem of guessing the missing elements of a distance matrix and consequently solved the problem by matrix factorization [6]. Hariri, *et al.*, proposed a decentralized coordinate-based solution for Internet delay measurement by modeling the latency between each pair of nodes as a virtual distance between those nodes in a virtual coordinate system [7]. There has also been some other work that is related to delay measurement [8, 9].

The above work, however, has mainly focused on the measurement and estimation of delay but has hardly touched the issue of correlation of nodes with respect to delays, not to mention a thorough study of correlation between delays of network nodes. On the other hand, understanding such correlation between network nodes can provide us with an insight on how to architect a measurement infrastructure for the Internet based on delays as well as on how to design new methods for the measurement and estimation of packet delays in the Internet. It can also help to address the issue of how to select or where to deploy network monitors throughout the Internet [10, 11, 12]. For example, for delay measurement, we can develop algorithms to systematically choose some nodes dynamically as the monitoring nodes to perform actual measurement while making some other nodes estimate delays based on the correlation of RTTs between these two groups of nodes.

In this paper, based on experiment data that we have obtained using OPNET, we study how network topological properties affect correlation. In particular, we analyze the effect of path length on the correlation of RTTs between network nodes and draw the following main conclusions for such correlation. First, correlations don’t exhibit significant difference for path length combinations that only differ in the length of the shorter private path but will become the weakest in statistics when the lengths of the two private paths become the same. Second, correlations are not much different for path length combinations that have the same path length ratio (which is defined as the ratio of the length of the common path to that of the longer private path). Third, the higher the path length ratio is, the stronger the correlation will get with the ratio 1/1 being the inflection point. These conclusions can be used to guide the design of algorithms for the estimation of delays with desired accuracy and for the deployment or selection of network nodes to perform delay measurement to achieve optimization in network measurement.

The rest of this paper is organized as follows. In Section 2, we briefly describe the experiment that we have performed to generate the data on which we rely to perform the correlation analysis. In Section 3, we describe the formula for characterizing correlation along with some notations that will be used throughout our discussion. In Section 4, we study how path lengths and path length combinations affect correlation to reach two major conclusions. In Section 5, we study how path length ratios affect correlation to reach the third major conclusion. In Section 6, we study the relationship between correlation and path length ratios to reach the fourth major conclusion. Finally, in Section 7, we conclude this paper in which we also discuss our future research directions.

2. The Experiment

Our goal in this analysis study is to quantify correlation between RTTs of two source nodes to a common destination node, in particular, the effect of path length on such correlation which describes to what extent the two sequences of RTT values that are observed respectively by the two source nodes have the same changing trend.

2.1. Experiment method and simulation topology

In our study, we first use PING to get RTT values over our experiment network and then the simulation tool OPNET to get RTT values with defined number of hops and level of load. The abstraction of our simulation topology is shown in Figure 1 in which we model the network with three subnets: two host the two source nodes that send PING packets for the measurement and the third hosts the destination node that receives the PING packets. In the subnets, the workstations involved send or receive PING packets and, at the same time, run the Raw Packet Generator (RPG) module offered by OPNET to generate self-similar traffic as the background traffic for the simulation network. Of course, there can be one or more other subnets between the ones that host the source and the destination nodes. We can see that, by varying the topological connections between the two source nodes and the common destination node in the simulation experiment using this abstraction, we can generate a full set of data that allows us to observe RTTs and to fully study the correlation of RTTs between the two source nodes to the common destination node. Such an abstraction captures all possible topological relationships between the two source nodes in terms of path lengths that can affect the correlation of RTTs.

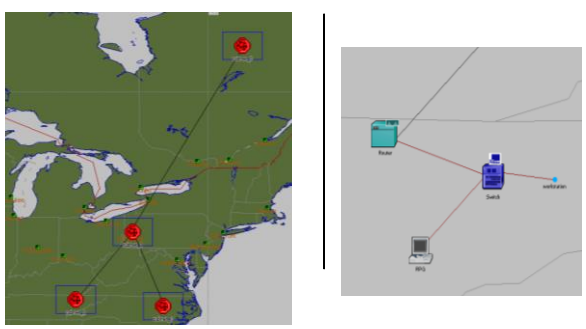


Figure 1. Abstraction of the simulation topology

2.2. Experiment results

We define the following three parameters in our experiment:

- Path length. As pointed out in [13], the average number of hops in the Internet is 15-19 and the default maximum length of any path in the OPNET is 15 hops. Therefore, in our experiment, we set the maximum length of any path to be 15 hops, which makes our experiment practical while still meaningful.

- Path load. We model the load for the network in 3 tiers in the experiment: light, with an average utility of each path being 20% within the range [15%-25%]; medium, with an average utility of each path being 50% within the range [40%-60%]; and heavy, with an average utility of each path being 80% within the range [65%-95%].
- Traffic pattern. The pattern for the Internet traffic has been shown to be self-similar rather than Poisson [14]. Therefore, we use the RPG module in OPNET to generate self-similar traffic with a typical Hurst value of about 0.7 [15].

Based on our simulation topology and experiment configuration, we performed 13,259 experiments using OPNET that cover all possible path lengths and loads and, as the result, have generated a set of 22,748,816 data for RTTs to allow us to perform a thorough study on the correlation of RTTs between two source nodes to a common destination node.

3. Formula and Notations

3.1. Correlation formula

In our study, we use rank-based Kendall's method to compute correlation [16] in which τ is defined as follows:

$$\tau = 1 - \frac{2s(\pi,\sigma)}{N(N-1)/2} \qquad (1)$$

Suppose Y = y_1, …, y_n is a set of items to be ranked. Then, in this formula, π and σ are two distinct orderings of Y, $s(\pi,\sigma)$ is the minimum number of adjacent transpositions that needs to be performed to bring π to σ, and N is the number of items to be ranked. It has been shown that Kendall's method is less sensitive to outliers and non-normality than the standard Pearson estimate [17].

3.2. Path lengths and loads

Let A and B be the two source nodes and T be the common destination node, as shown in Figure 2. It is clear that Paths A-T and B-T share a common path M-T with length H_{com} and load on the path is U_{com}. Similarly, the length of private path A-M is H_{pri_a} and load on the path is U_{pri_a}. Same can be said for node B, *i.e.*, B-M, H_{pri_b} and U_{pri_b}. In our analysis study, $H_{com}+\max(H_{pri\_a},H_{pri\_b})\leq 15$ and $0\leq U_{com}$, U_{pri_a}, $U_{pri\_b}\leq 100\%$. We call "H_{com}-H_{pri_a}-H_{pri_b}" a path length combination (LEC for short) and "U_{com}-U_{pri_a}-U_{pri_b}%" a path load combination (LOC for short).

In our analysis study, we mainly examine the cases in which the length of the common path $H_{com}\geq 1$ and so is the length of the shorter private path. This is because when $H_{com} = 0$, little remains in common for nodes A and B in terms of topological connections to T, making their RTTs in this case hardly correlated. We can thus derive a total number of 417 length combinations in the analysis. For network loads, since we use three representative tiers in the analysis, *i.e.*, 20%, 50% and 80% to represent light, medium and heavy load, respectively, we can thus derive a total number of 27 load combinations in the analysis.

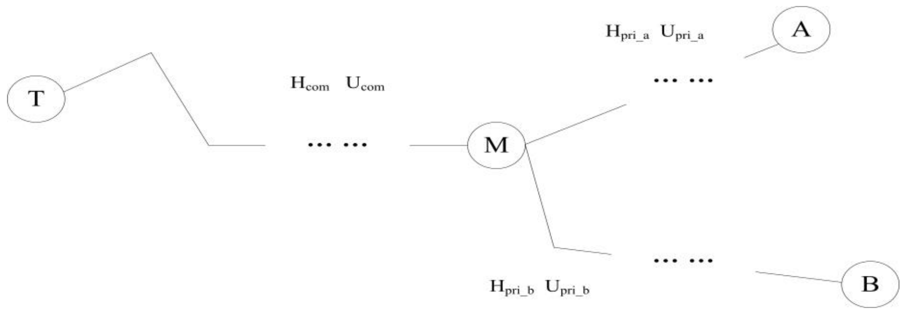


Figure 2. Notations for path lengths and path loads

4. Correlation Analysis for Path Lengths

We study how the lengths of the three paths, *i.e.*, the common path and the two private paths, as well as path length combinations affect correlation in this section and reach some major conclusions regarding correlation of RTTs.

There are always three paths in a path length combination and changing the length of any of them results in a new path length combination. Even after we fix the lengths of the common and the longer private paths, there could still be one or more path length combination resulting from one or more different lengths for the shorter private path. For each of the 27 load combinations, we measure the RTTs and derive the corresponding correlations using Formula (1) for different path length combinations and then analyze the results.

4.1. Effective private path lengths

Let's first see how the length of the longer private path can have any meaningful effect on correlation. Figure 3 shows the results of correlation as the length of the longer private path varies from 14 to 1 (from 1 to 14). Note in the figure that almost all the data for the same path length from the experiment fall into a single range and the bold line within the range is the median of the results. It is clear from the figure that when the length of the longer private path is 8 hops or more, correlation is almost always below 0.2, a level that we consider to be too weak to be useful in actual RTT measurement and estimation and, consequently, to be worthy of further analysis. Therefore, we can use these results to reduce the amount of analysis work so that we can focus the analysis on more useful path length combinations. Moreover, even less or some more analysis work can be performed based on the results shown in Figure 3.

4.2. The effect of shorter private path lengths

The results of RTT correlations for the cases in which only the length of the shorter private path varies while the lengths of the common and the longer private paths remain unchanged are shown in Figure 4 in which, for clarity, we only show the results for seven different common path lengths, for the results for other common path lengths follow essentially the same pattern. Since the length of any complete path is at most 15 hops in our analysis, when the length of the longer private path is 7 hops, the length of the common path can be 8 hops at the most. Consequently, the length of the shorter private path can be between 1 and 7 hops, allowing us to present as many as 7 results for each common path length. For example, when the lengths of the common and the longer private paths are 9 and 6 hops, respectively, the length of the shorter private path can vary from 1 to 6 hops, having as many as 6 results for this particular common and longer private path length combination. We perform the analysis for all possible common and longer private path length combinations and the results are the average of those for all possible load combinations for the same path length combination so as to provide a complete picture on the effect of path length on the correlation.

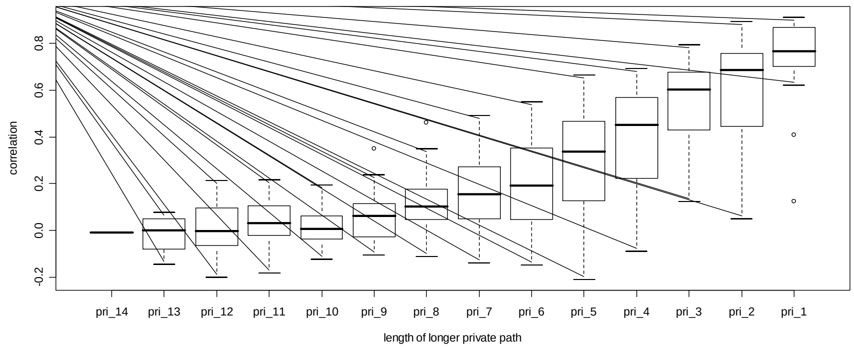


Figure 3. Correlation as the length of the longer private path varies

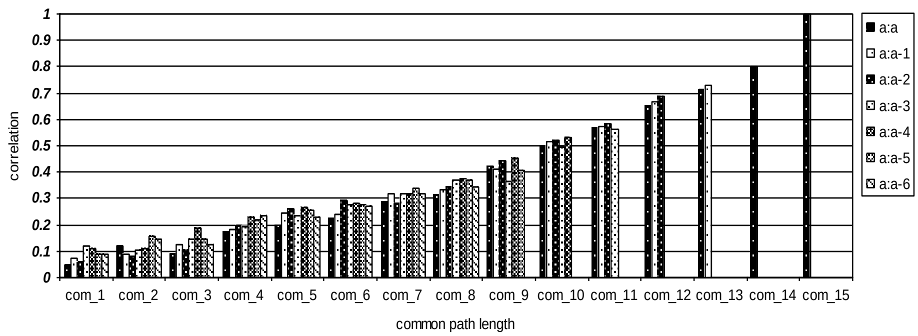


Figure 4. Correlation as the length of the shorter private path varies

We can see from Figure 4 that correlations are not much different for the length combinations in which the lengths of the common and the longer private paths remain unchanged. That is, the length of the shorter private path doesn't affect the correlation too much and it is the lengths of the other two paths that play a more significant role on the correlation. We perform more formal analysis below to explain this phenomenon.

4.3. Variance analysis

We use a mathematical method based on variance analysis to support the results illustrated in Figure 4 in which we show that varying lengths of the shorter private path won't cause significant difference to correlation.

For variance analysis, we arrange the correlations in a three-dimensional matrix in which the first dimension contains the 7 different private path length combinations denoted as a:a, a:a-1, …, a:a-6 as shown in Figure 4. Note that there is no private path length combination when the length of the common path is 15 hops. In this case, the value for the correlation is simply 0. The second dimension in this matrix contains the 14 different lengths for the common path and the third dimension contains the 27 load combinations.

From this three-dimensional matrix, we extract 14 two-dimensional arrays in each of which the rows represent the 27 load combinations and the columns the number of private path length combinations that have non-zero values for the correlation. We then do the variance analysis for each of the 14 arrays. When the length of the common path varies from 1 to 12 hops, the corresponding arrays will have 3 or more columns, so we do the analysis with Friedman test [18]. Let R_{ij} be the rank of x_{ij} within block i. The test statistic for Friedman test can be carried out using the following formulas:

$$Q = \frac{12N}{s(s+1)} \sum_{i=1}^{s} (R_{i\cdot} - \frac{1}{2}(s+1))^2 \tag{2}$$

$$R_{i\cdot} = \frac{1}{N}(R_{i1} + R_{i2} + \ldots + R_{iN}),\ i = 1,2,\ldots,s \tag{3}$$

When the length of the common path is 13 hops, the corresponding arrays will only have 2 columns, so we do the analysis with Binom test [18] which is more suitable for verifying whether two sets are significantly different. Let p be the probability for getting the first category and q=1–p be that for getting the other category. The formula for Binomial test of significance is as follows:

$$p(r)_{binomial} = {}_nC_r * p^r * q^{n-r} = \frac{(n!\,p^r q^{n-r})}{(r!(n-r)!)} \tag{4}$$

We don't do any analysis when the length of the common path becomes 14 hops, for the corresponding array will only have 1 column.

The threshold value for the variance analysis for both scenarios is 0.05 [18] and the results are shown in Figure 5.

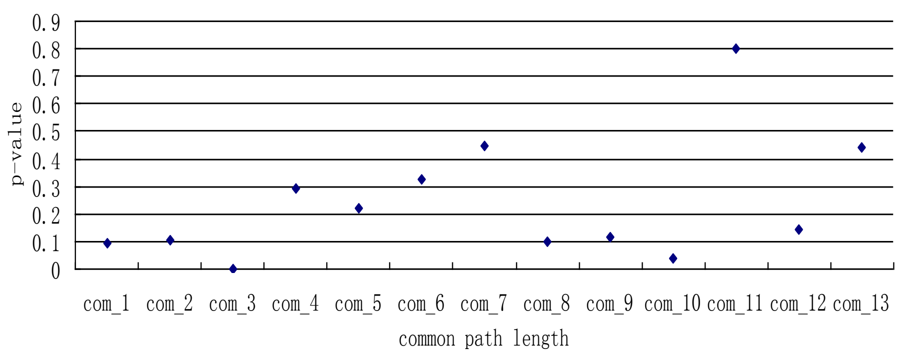


Figure 5. P-values of variance analysis for the shorter private path length

We can see from Figure 5 that almost all the p-values are far above the threshold of 0.05. We can thus conclude that differences in the length of the shorter private path won't cause much difference to the correlation. In another word, the length of the shorter private path has an insignificant effect on correlation. Therefore, for any load combination, the common and the longer private path lengths together play the dominating role on correlation of RTTs regardless of the length of the shorter private path.

4.4. The effect of longer private path lengths

Although variance analysis shows that there is no significant difference among correlations for path length combinations in which only the length of the shorter private path varies, changing the length of the longer private path could have some effect on correlation. We now perform analysis on the effect of the longer private path length on correlation to identify representative path length combinations.

We define Set_{CLL} as the set of all length combinations in which the lengths of the common and the longer private paths are the same, that is:

$$LEC(x_i, y_i, z_i) \in Set_{CLL}(c, l) \text{ iff } x_i=c \text{ and } \max(y_i, z_i)=l,\ l \leq i \leq |Set_{LEC}| \quad (5)$$

Note that $|Set_{LEC}|$ is the total number of such length combinations. With this definition, we apply further analysis on the arrays introduced in 4.3. Here, we extract 27 two-dimensional arrays, each corresponding to one of the 27 load combinations. The rows of each array are the private length combinations in which the number of non-zero value items range from 1 to 7 depending on the length of the corresponding common path and the columns are the 13 common path lengths. Again, we are only interested in common path lengths of up to 13 hops, for there is only one correlation value when the common path length is 14 hops and, consequently, no comparison can be made.

In each of the 27 two-dimensional arrays, for every common path length, we arrange the correlation values in the column in an ascending order and assign a value (starting from 1) to a correlation that corresponds to its place in the ordering. There are thus at most 13 values based on such orderings for any private path length combination. We can then form 27 two-dimensional arrays corresponding to the above 27 arrays in which correlation values are replaced with ordering values. We further average out all the ordering values from the 27 arrays for the same array entry and list the results in Table 1 with the last column being the mean of each row, *i.e.*, the average of the values for a private path length combination over all the 13 common path lengths. We can see from the table that private path length combination “a:a” has the smallest mean value, which corresponds to the weakest correlation. We can thus conclude that, for any given common path length, among all private path length combinations with the same longer path length, correlation is the weakest when the length of the shorter private path becomes the same as that of the longer private path.

4.4. Summary

From the above analysis, we can reach the following conclusions for any load combination:

- RTT correlation is determined primarily by the length of the common path. For instance, when the length of the common path is 10 hops or longer (*i.e.*, the length of the private paths is 5 hops or shorter), correlation reaches 50% or higher (see Figure 4).
- When the lengths of the common and the longer private paths remain unchanged, varying the length of the shorter private path won’t cause significant change in correlation. Moreover, correlation becomes the weakest in statistics when the lengths of the two private paths become the same.

5. Correlation Analysis for Path Length Ratios

We have examined correlation between RTTs of two sources nodes for different common path lengths as well as for different private path length combinations. Based on the conclusions we have reached, we now examine correlation for more dynamic path configurations, *i.e.*, for path length combinations with the same path length ratio which is defined as the ratio of the length of the common path to that of the longer private path.

Table 1. Mean values for private path length combinations

	com_1	com _2	com _3	com _4	com _5	com _6	com _7
a:a	3.962963	3.962963	3.370370	3.407407	3.185185	3.518519	3.555556
a:a-1	4.111111	3.592593	3.518519	3.814815	3.777778	3.444444	4.148148
a:a-2	3.296296	3.814815	3.518519	3.666667	4.185185	4.333333	3.444444
a:a-3	3.888889	3.814815	4.148148	3.592593	4.185185	4.185185	4.074074
a:a-4	4.222222	4.185185	4.962963	5.111111	4.259259	3.925926	4.148148
a:a-5	3.814815	4.592593	4.481481	4.185185	4.370370	4.296296	4.629630
a:a-6	4.185185	3.777778	4.000000	4.222222	4.037037	4.296296	4.000000
	com _8	com _9	com _10	com _11	com _12	com _13	mean
a:a	3.185185	3.555556	2.333333	2.296296	1.703704	1.370370	3.031339
a:a-1	3.592593	3.000000	2.888889	2.555556	2.074074	1.629630	3.242165
a:a-2	3.777778	3.888889	3.444444	2.629630	2.222222	X	3.518519
a:a-3	4.851852	2.925926	2.814815	2.518519	X	X	3.727273
a:a-4	4.037037	4.074074	3.518519	X	X	X	4.244444
a:a-5	4.333333	3.333333	X	X	X	X	4.243802
a:a-6	4.222222	X	X	X	X	X	4.150235

5.1. Analysis of correlations

In our analysis, we express path length combinations as (x, y, z_1) and (n*x, n*y, z_2) where x, y and z_1 and z_2 are the lengths of the common, the longer private and the shorter private paths, respectively, and define path length ratio r below:

$$r = \frac{H_{com}}{\max(H_{pri\_a}, H_{pri\_b})}, \; 1 \le H_{com}, H_{pri\_a}, H_{pri\_b} \le 14 \qquad (6)$$

Thus, r is the ratio of the length of the common path to that of the longer private path. Since the length of any path is at most 15 hops, we can deduce that there is a total number of 71 path length ratios. Let's further define path length ratio set R_r to be a set that includes all the path length combinations that have the same path length ratio, that is,

$$\mathrm{LEC}(x_i, y_i, z_i) \in R_r \text{ iff } x_i/y_i = n, \; 1 \le i \le |Set_{LEC}| \qquad (7)$$

From the experiment data, correlations for path length combinations that have the same path length ratio can be determined. For purpose of clarity, we only show the results that correspond to integer path length ratios. Since the length of any complete path is at most 15 hops, there are at most 11 integer path length ratios, from 1/6 to 6/1, as shown in Figure 6 in which "x to y" denotes a path length ratio and "nc" denotes a path length combination within a path length ratio set. We can then see that correlations are very close to each other for path length combinations that have the same path length ratio, especially when the path length ratio is 1/1 or lower. Our reasoning is that correlations are already very weak when the path length ratio is 1/1 or lower. Even as the number of path length combinations increases, the correlations won't change too much. When the path length ratio is higher than 1/1, however, correlations increase rather quickly as the value of the ratio increases. But for path length combinations that have the same path length ratio, correlations still remain close to each other.

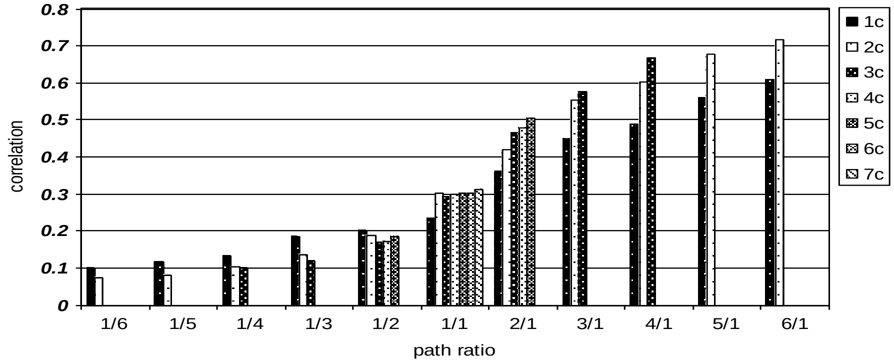
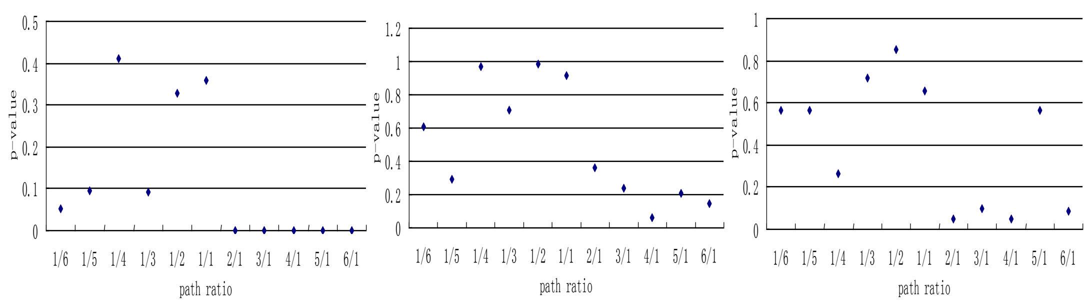


Figure 6. Correlations for path length combinations with different path length ratios

5.2. Variance analysis

We perform variance analysis again to verify the results shown in Figure 6. First, we construct a three-dimensional matrix of 11*27*7 in which the first dimension denotes the 11 path length ratios, the second denotes the 27 load combinations and the third denotes the maximum number of cases for each path length ratio which is 7. Second, we extract from the three-dimensional matrix 11 two-dimensional arrays each of which has 27 rows that correspond to the 27 load combinations and 2 to 7 columns that correspond to the number of cases for the respective path ratio. Third, we perform variance analysis for each two-dimensional array using Binom test when there are only 2 cases for the path ratio and Friedman test when there are 3 or more. The results of the p-value from the analysis are shown in Figure 7(a) from which we can see that all the p-values are beyond the threshold 0.05 (the p-value for path length ratio of 1/6 is 0.0508) when the path length ratio is 1/1 or lower. But for path length ratios higher than 1/1, all the p-values are below the threshold 0.05. So, the number of path length combinations for various path length ratios has non-trivial effect on correlation when path length ratio is higher than 1/1. When the ratio is 1/1 or lower, correlations for path length combinations that have the same path length ratios will only show insignificant differences to each other.



(a) (b) (c)

Figure 7. Distrbution of the p-values for different path length ratios

Since the results for path length combinations with path length ratios of higher than 1/1 don't look as nice as those with path length ratios of 1/1 or lower, we perform a finer analysis on set R_r using Wilcox test [18] when there are only 2 path length combinations for a path ratio and Kruskal test [18] when there are 3 or more to check if the median correlation values have much difference. We apply this method to check the differences among the median correlation values for different path length combinations in the same path length ratio set.

Suppose there are two vectors X and Y that have n_1 and n_2 items, respectively. We introduce a new vector Z to include all the items in X and Y and order the items in this new vector. Then, every item has its own rank based on its position in the new ordering and we denote the rank of the items from Y as R_i. The test statistics W for Wilcox test are calculated using the following formula:

$$W = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i=1}^{n_2} R_i \tag{8}$$

For $N=n_1+n_2+\ldots+n_g$ measurements that are based on g methods each of which has a group of n_i items, denote r_{ij} as the rank of the jth item in group i ($1 \le i \le n_g$) in the total ordering of the N items. Then, the test statistics K for Kruskal test are calculated using the following formulas:

$$K = \frac{12}{N(N+1)} \sum_{i=1}^{g} n_i (\overline{r_{i\cdot}} - \frac{N+1}{2})^2 \tag{9}$$

$$\overline{r_{i\cdot}} = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}\ ;\ \bar{r} = \frac{1}{2}(N+1) \tag{10}$$

We can then compare every two path length combinations in the same path length ratio set. The results of the p-values based on Kruskal test are shown in Figure 7(b) from which we can see that nearly all the p-values of Kruskal test are far beyond the threshold 0.05 (actually, the minimum is 0.05846765). Therefore, we can conclude that the number of path length combinations for any path length ratio has trivial effect on the correlation and the median correlation values for the path length combinations in the same path length ratio set have insignificant difference. We can then use a median correlation value for a path length combination in set R_r to represent correlations for all the other path length combinations in the same set.

Moreover, when applying load combinations of 20-20-20%, 50-50-50% and 80-80-80%, we are able to identify a supposed-to-be more obvious property for the path length ratios. In the analysis, we still use Binom test when there are only 2 path length combinations in a path length ratio set and Friedman test when there are 3 or more to do the variance analysis. The p-values out of the analysis are shown in Figure 7(c) from which we can see that all the p-values are beyond the threshold 0.05 except one (the median for path length ratio 2/1 is 0.04824061). So, we can conclude that the correlations for different path length combinations with the same path length ratio have insignificant differences to each other regardless of network traffic condition.

5.3. Summary

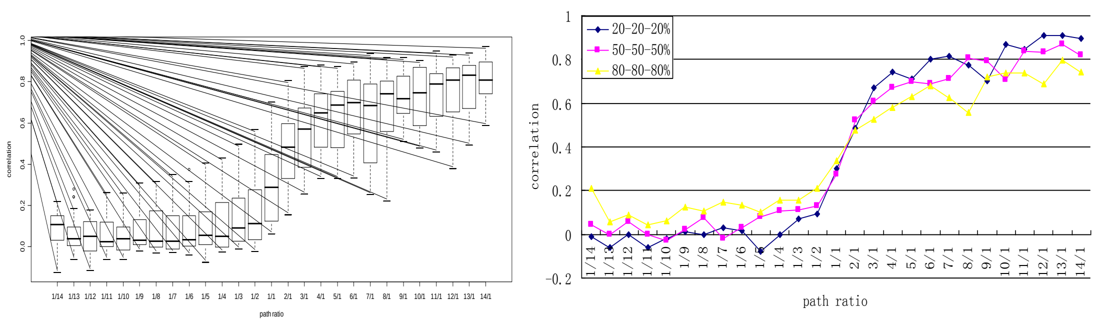
From the above analysis, we can reach the following conclusions for any load combination:

- The number of path length combinations with the same path length ratio has non-trivial effect on correlation when the path length ratio is higher than 1/1. Otherwise, correlations for path length combinations with the same path ratio have insignificant differences.

- Path length combinations with the same path length ratio exhibit trivial effect on correlation and the median correlation values for different path length combinations with the same path length ratio have insignificant differences. We can thus use the median correlation value for a path length combination to represent the median correlation values for all the path length combinations that have the same path length ratio.
- Correlations for different path length combinations with the same path length ratio won't be significantly different.

6. Relationship between Correlations and Path Length Ratios

Based on the analysis study that we have performed so far, we can now put things together to show the relationship between correlations and path length combinations. Figure 8(a) shows the distribution of correlations for every path length ratio statistically over a range with the median value being indicated with a bold line in the range. The median correlation value for a path length combination can then be used to represent correlations for different path length combinations with the same path length ratio. We can thus see that correlation exhibits an upward trend as path length ratio increases with 1/1 as the reflection point. We can also see that when path length ratio goes down to 1/6 or lower, correlation values fall to 0.37 or lower (within the range [-0.02073, 0.373886]), which can be considered weak, and, when path length ratio increases to 9/1 or higher, correlation values become 0.5 or higher (within the range [0.51734, 0.972678]), which can be considered strong.



(a) (b)

Figure 8. Relationship between correlations and path length ratios

For load combinations of 20-20-20%, 50-50-50% and 80-80-80% that are used to characterize general network traffic conditions, based on the results and conclusions reached so far, we can express correlations for path length combinations of the form "n*c,n*a,(a-m)" where "n" is a positive integer and "m" is also a positive integer less than a. The relationship between correlations and path length combinations can then be expressed using the relationship between representative correlations and path length ratios. Such a relationship is depicted in Figure 8(b) which shows that the higher the path length ratio is, the stronger the correlation becomes and that the path length ratio 1/1 is the inflection point. We can thus conclude that correlation of RTTs generally exhibits an upward trend and will get stronger along with the increase in path length ratio under any network traffic conditions.

7. Conclusion

In this paper, we presented a quantitative analysis on the effect of path length on the correlation of RTTs between two different source nodes to a common destination node using data generated with OPNET, from which we reached the following conclusions: (1) correlations have insignificant difference to each other for path length combinations that only differ in the length of the shorter private path, moreover, correlations become the weakest when the lengths of the two private paths become the same; (2) correlations won't have much difference to each other for path length combinations that have the same path length ratio; and (3) the relationship between correlations and path length ratios exhibits an upward trend and the higher the path length ratio is, the stronger the correlation becomes with the path length ratio of 1/1 being the inflection point, and, in particular, path length ratio of 1/6 or lower shows weak correlation while that of 9/1 or higher indicates strong correlation.

In the future, we will extend our work into real networks to verify the results and to further refine our analysis. We will also apply our results to the design of algorithms for RTT measurement and estimation and for the deployment or selection of measurement nodes for RTT measurement to further our research on network measurement.

Acknowledgments

The work presented in this paper has been supported by funding from National Natural Science Foundation of China (61272500), from Beijing Education Commission Science and Technology Fund (KM201010005027), and from Ministry of Science and Technology of China Public Sector Special Research Fund (201310040-03).

References

[1] http://www.opnet.com/.
[2] P. Francis, S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavitt and L. Zhang, "IDMaps: A global internet host distance estimation service", IEEE/ACM Trans. Networking, vol. 9, no. 5, **(2001)**.
[3] Y. Chen, K. H. Lim, R. H. Katz and C. Overton, "On the stability of network distance estimation", ACM SIGMETRICS Performance Evaluation Review, vol. 30, no. 2, **(2002)**.
[4] S. Agarwal and J. R. Lorch, "Matchmaking for online games and other latency-sensitive P2P systems", ACM SIGCOMM Computer Communication Review, vol. 39, no. 4, **(2009)**.
[5] Y. Chen, Y. Xiong, X. Shi, J. Zhu, B. Deng and X. Li, "Pharos: Accurate and decentralised network coordinate system coordinate system", IET Communications, vol. 3, no. 4, **(2009)**.
[6] Y. J. Liao, G. Pierre and L. Guy, "Network distance prediction based on decentralized matrix factorization networking", Proceedings of the 9th International IFIP-TC 6 Networking Conference, **(2010)** May 11-15; Chennai, India.
[7] N. Hariri, B. Hariri and S. Shirmohammadi, "A distributed measurement scheme for Internet latency prediction", IEEE Trans. Instrumentation and Measurement, vol. 60, no. 5, **(2011)**.
[8] K. Peter, M. Laurent and T. Dan-Cristian, "Non-metric coordinates for predicting network proximity", Proceedings of the 27th IEEE Conference on Computer Communications, **(2008)** April 14-17; Phoenix, AZ, USA.
[9] G. F. Wu, Q. Dou and H. Q. Fu, "Passive-landmark based network distance prediction", Proceedings of International Forum on Information Technology and Applications, **(2009)** May 15-17; Chengdu, China.
[10] D. Shlomi, E. Yuval and P. Rami, "Incremental deployment of network monitors based on Group Betweenness Centrality", Information Processing Letters, vol. 109, no. 20, **(2009)**.
[11] H. H. Shi, J. S. He and X. Xu, "A network traffic monitor-node selection algrothim based on triple tag", Journal of Beijing University of Posts and Telecommunications, vol. 32, **(2009)**.
[12] L. Yan and C. Marek, "Approximation algorithms for the fault-tolerant facility placement", Information Processing Letters, vol. 111, no. 11, **(2011)**.
[13] A. Fei, G. Pei, R. Liu and L. Zhang, "Measurements on delay and hop count of the Internet", Proceedings of IEEE Global Telecommunications Conference, **(1998)** November 8-12; Sydney, Australia.
[14] W. E. Legend and M. S. Taqqu, "On the self-similar nature of Ethernet traffic", IEEE/ACM Trans. Networking, vol. 2, no. 1, **(1994)**.
[15] P. Leys, J. Potemans, B. Van den Broeck, J. Theunis, E. Van Lil and A. Van de Capelle, "Use of the raw packet generator in OPNET", Proceedings of OPNETWORK 2002, **(2002)** August 28; Washington, DC, USA.

[16] M. Lapata, "Automatic evaluation of information ordering: Kendall's tau", Computational Linguistics, vol. 32, no. 4, **(2006)**.
[17] K. C. Lan and J. Heidemann, "A measurement study of correlations of Internet flow characteristics", Computer Networks, vol. 50, no. 1, **(2006)**.
[18] Y. Xue, "Statistical Modeling and R Software", 2nd Edition, Tstinghua University Press, Beijing, China **(2008)**.

Authors

Nafei Zhu received a Bechalor's degree in Communication Engineering and a Master's degree in Traffic Information Engineering & Control, both from Central South University in Changsha, China, and a Ph.D. degree in Computer Science and Technology from Beijing University of Technology in Beijing, China. She is currently an engineer in the National Application Software Testing Labs in Beijing, China. Ms. Zhu's main research interests include measurement and analysis of network performance and security and she has published several research papers in scholarly journals and international conferences in the above research areas.

Jingsha He received his B.S. degree in Computer Science from Xi'an Jiaotong University in Xi'an, China and his M.S. and Ph.D. degrees in Computer Engineering from University of Maryland at College Park in U.S.A. He is currently a professor in the School of Software Engineering at Beijing University of Technology in Beijing, China. Professor He has published over 170 research papers in scholarly journals and international conferences and has been granted nearly 30 patents in the United States and in China. Prof. He's main research interests include information security, network measurement, and wireless ad hoc, mesh and sensor network security.

Yue Zhou is currently the Director of the National Application Software Testing Labs in Beijing, China. She has over 10 years of experience in software development and testing. Ms. Zhou is an expert in software quality assurance and has participated in many national projects and software quality assurance programs such as the Beijing Olympic Games in 2008 and the Shanghai World Expo in 2010. She has also co-authored several national standards in areas related to software quality assurance. Ms. Zhou's main research interests include the application of information technologies and software quality assurance.

Wei Wang received her Bachelor's degree in Computer Science and Technology from Tsinghua University in Beijing, China. She is currently a department manager in the National Application Software Testing Labs in Beijing, China. Ms. Wang has participated in over 100 projects in finance, gaming, petroleum, taxation, irrigation, accounting and public health. She has also taken part in drafting several national standards in the above areas.