Master's Thesis

Title

A study on reduction of inter-ISP transit cost caused by overlay routing

Supervisor

Professor Hirotaka Nakano

Author

Kazuhito Matsuda

February 15th, 2010

Department of Information Networking

Graduate School of Information Science and Technology

Osaka University

Master's Thesis

A study on reduction of inter-ISP transit cost caused by overlay routing

Kazuhito Matsuda

Abstract

Overlay routing is an application-level routing mechanism on overlay networks. Existing researches on overlay routing revealed that it can improve user-perceived performance by choosing a path based on network performance metrics such as end-to-end latency and available bandwidth. On the other hand, overlay routing may harm ISPs' cost structure because of the policy mismatch between IP routing and overlay routing. IP routing provided by ISPs is generally configured based on the monetary cost of links interconnecting to neighboring ISPs. On the other hand, the overlay routing selects the path based on end-to-end network performance. Consequently, the overlay routing may utilize paths which traverse additional transit links between ISPs, which requires additional inter-ISP transit cost. One possible solution to this problem is to limit the overlay routing not to utilize paths which increase inter-ISP transit cost largely.

In this thesis, the author proposes a method to reduce inter-ISP transit cost caused by overlay routing. For this purpose, the number of inter-ISP transit links on a path is utilized as a metric of transit cost and the overlay path is chosen to decrease the transit cost while keeping the effectiveness of overlay routing itself. Since there is no public information on the number of transit links on a path, and since it cannot be obtained by simple end-to-end measurement methods, the author builds up a method to estimate the number of transit links on a path from end-to-end network performance values which can be measured easily by overlay nodes. Through the multiple regression analysis of end-to-end network performance values, the regression equation to estimate the number of transit links is obtained.

To confirm the effectiveness of the proposed method, the author first assesses the estimation accuracy of the regression equation and then evaluates the performance of limited overlay routing which suppresses the number of transit links determined by the proposed estimation method. Through extensive evaluations using measurement results in the actual network environment, the author confirms that the absolute underestimation error in the number of transit links on overlay-level paths is smaller than one in almost 80% of overlay paths, and that the overlay routing with the proposed method can achieve almost the same degree of improvement as that without limitation, while decreasing the number of the transit links traversed by overlay-routed paths.

Keywords

overlay network
overlay routing
inter-ISP transit cost
multiple regression analysis
PlanetLab

Contents

List of Figures

List of Tables

1 Introduction

Overlay network is defined as an application-level logical network built upon IP networks, and it provides services including what the traditional IP network cannot provide, such as Quality of Services (QoS) improving [1, 2] and multicast routing [3, 4]. Overlay routing is an application-level routing mechanism on overlay networks [5-9], which provides application-level routes for network application traffic, as depicted in Figure 1. One early and typical example is the Resilient Overlay Network (RON) [10], in which each overlay node measures the end-to-end latency and packet loss ratio of the network paths to other nodes, and determines the path for the overlay network traffic, originating from the node, which can be either a direct path from the node to the destination node or a relay path that traverse other node(s) before reaching the destination node (Figure 1). In this thesis, the terms "overlay routing" and "IP routing" are utilized to refer to traffic routing at the application level and the IP level, respectively.

One of advantage of overlay routing is that user-perceived performance can be improved compared with IP routing [3, 7, 11-15]. Most of these studies focused on evaluations only with end-to-end latency as an overlay routing metric. In [16], our research group revealed that a significant performance improvement can be obtained by overlay routing based on bandwidth-related information such as available bandwidth and TCP throughput. Such performance improvement is caused primarily by the policy mismatch between IP routing and overlay routing. IP routing is mainly based on router-level hop count and Autonomous System (AS) level hop count. Besides, Internet Service Providers (ISPs) who operate IP routing make their routing decisions based on the monetary contracts against their neighboring ISPs, which are either transit or peering relationships. Transit links and peering links has different monetary charge mechanisms, and each ISP utilizes the routing configurations affected by such differences. On the other hand, overlay routing primarily chooses the paths to enhance user-perceived performance such as end-to-end latency and available bandwidth.

Whereas this policy mismatch would bring the improvement of end-to-end network performance, it also generates a problem for ISPs' cost structure and increase the inter-ISP transit cost for the entire network [17, 18]. To reduce inter-ISP transit cost, locality aware methods have proposed [19-21]. However, it is difficult to obtain locality information, that means which ISP each node is in, and it may decrease user-perceived performance provided by overlay routing.

One possible solution to this problem is to limit the overlay routing not to utilize paths which increase the inter-ISP transit cost largely. However, the contract information between ASes (ISPs), which determine that whether the relationship between ASes is transit or peering, is not available directly and there is no simple end-to-end measurement method to obtain the number of transit links. Meanwhile, our research group has revealed in [16] that the inter-ISP transit cost on the end-to-end path may be estimated by network performance values measured easily by overlay nodes. However, [16] presented no practical method to reduce inter-ISP transit cost in overlay routing.

In this thesis, the author proposes a novel method to decrease the inter-ISP transit cost caused by overlay routing, while keeping the performance improvement provided by overlay routing itself. The author defines a metric for the inter-ISP transit cost on a path chosen by overlay routing as the number of transit links on the path and proposes a limited overlay routing with the metric. To calculate the metric value, the author builds up a method to estimate the number of transit links on a path from end-to-end network performance values which can be measured easily by overlay nodes such as router-level hop count, end-to-end latency, and available bandwidth. By using multiple regression analysis of network performance values, the regression equation is derived to estimate the number of transit links on a path.

The author confirms the effectiveness of the proposed method by evaluating the performance of the overlay routing which is assumed to be operated on the overlay network on PlanetLab environment and Japanese commercial network environment. First, to set a baseline for the discussion, the author evaluates the performance improvement of the overlay routing without transit link limitation. The author then evaluates the limited overlay routing which has the limitation on transit links by using the true information on inter-ISP relationships. The author also evaluates the performance of limited overlay routing which suppresses the utilization of transit links by the proposed estimation method.

The rest of this thesis is organized as follows. In Section 2, the research background on the overlay routing and its problem to the inter-ISP transit cost are described. In Section 3, the author introduces the metrics for the overlay routing to choose a path. The author also presents a metric of the inter-ISP transit cost and proposes a method to reduce inter-ISP transit cost with the metric. In Section 4, how to obtain the dataset used in this thesis is described. In Section 5, the author calculates the correlations between the number of transit links on a path and network

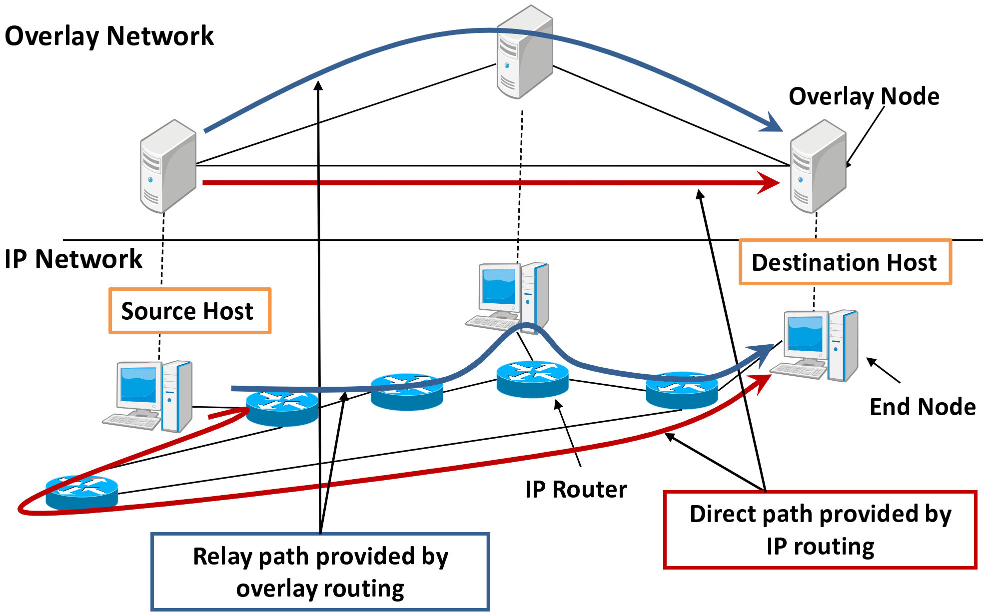


Figure 1: Overlay routing

performance values and derives the regression equation. Then the author evaluates the accuracy of the regression equation. Besides, the author investigates the difference in the network properties in the PlanetLab environment and the Japanese commercial network environment by using the analysis results. In Section 6, the results of numerical evaluation are presented to confirm the effectiveness of the proposed method. Finally, in Section 7, the author summarizes the conclusion and discusses future work.

2 Overlay routing and its problem

2.1 Overlay routing

Overlay routing is a technique for network application to improve end-to-end network performance by choosing the paths based on application-level network performance metrics such as end-to-end latency, available bandwidth, and TCP throughput.

On the other hand, IP routing is mainly based on metrics such as router-level and AS-level hop count, which do not always correlate with user-perceived performance. In addition, ISPs who operate IP routing have their own cost structure based on commercial contracts with neighboring ISPs and the routing configurations are largely affected by the cost structure. As mentioned in Section 1, there are two types of link between ISPs[1]. One is a transit link that connects the upper-level ISP and the lower-level ISP, and the other is a peering link used for peering relationship. The monetary cost of a transit link is usually determined by the amount of traffic traversing the link and the transit link is allowed to be used by the customers belonging to the interconnected ISPs. On the other hand, there is almost no monetary charge for a peering link, except for the cost paid to carrier companies for the physical link facilities. Therefore the peering link is allowed to be used only by the traffic whose origin and destination are the interconnected ISPs. ISPs make routing decisions by considering such differences between transit and peering links.

The advantage of overlay routing is mainly a result of the policy mismatch between IP routing and overlay routing mentioned above. Figure 1 shows a typical example of the advantage. We assume the IP routing uses direct path which is represented by the red arrow in the figure and the overlay routing chooses the relay path which is represented by the blue arrow. The arrow length in the IP network represents the end-to-end latency. When we compare the IP routing and the overlay routing from the source host to the destination host in this figure, the direct path provided by the IP routing has smaller router-level hop counts, but longer end-to-end latency comparing the relay path provided by the overlay routing. Therefore, the overlay routing provides better user-perceived performance (i.e., end-to-end latency) than the IP routing.

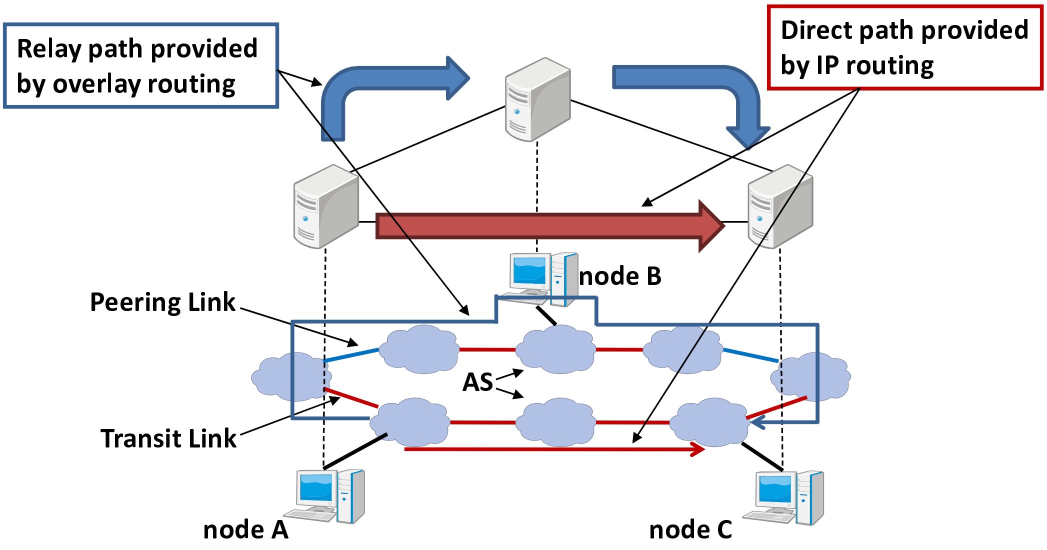


Figure 2: Increase of number of transit links by overlay routing

2.2 Influence on the cost structure of ISPs

Although overlay routing can improve user-perceived performance as described above, overlay routing may also generate traffic that does not follow the cost structure of the ISPs (i.e., the policy of IP routing provided by the ISPs), and the ISPs may incur additional monetary cost caused by such traffic. By accumulating such cost increase, it would increase the inter-ISP transit cost in the entire network.

Figure 2 shows a simple example of the problem. There are three endhosts and all of them work as overlay nodes. There are three overlay links between Node A and Node B, Node B and Node C, Node C and Node A. Each overlay link includes some underlay (i.e., IP) links and each underlay link is either a transit link or a peering link. In Figure 2, red links between ASes represent transit links and blue links between ASes represent peering links.

It is assumed that Node A generates traffic to Node C. When using the IP routing or the overlay routing which chooses the direct path (the red arrow in the figure), the traffic traverses two transit links. On the other hand, when the overlay routing uses the relay path via Node B (the blue arrow), the traffic traverses the transit links between Node A and Node B, and those between Node B and Node C. So the sum of transit links traversed by the relay path increases by two comparing with

[1]We can ignore sibling links because they connect ASes which belong to the same organization.

the direct path. As a consequence, the inter-ISP transit cost in the entire network increases.

In this thesis, in an attempt to resolve the above-mentioned problem, the author evaluates the overlay routing mechanism focusing on the problem and proposes a novel method to reduce the degree of the problem while keeping user perceived performance.

3 Metrics for overlay routing

In this section, the metrics for the overlay routing to choose overlay-level paths are explained. The author first introduces the metrics for the overlay routing to choose a path and also presents a metric of inter-ISP transit cost. Then, the author proposes a method to reduce the inter-ISP transit cost with the metric.

In this thesis, there are two candidates of overlay path, as described below.

- **direct path** - A direct path from the source node to the destination node, i.e., a one-hop path with overlay routing.
- **relay path** - A path from the source node to the destination node via another node. In this thesis, the author considers only the two-hop path, because paths with more than two hops do not contribute to improve user-perceived performance [16].

3.1 End-to-end latency

The overlay routing based on end-to-end latency is suitable for applications that require quick response rather than longer-term throughput such as Skype [22] and VNC [23]. The end-to-end latency of the overlay link between overlay nodes i and j is denoted as δ_{ij}. For the end-to-end latency of a relay path, the sum of the latencies of the overlay links that make up the relay path is utilized. Then, we determine the end-to-end latency of the direct path between nodes i and j, which is denoted as D^1_{ij}, and end-to-end latency of the relay path via node k, D^2_{ijk}, respectively, as follows.

$$D^1_{ij} = \delta_{ij} \tag{1}$$

$$D^2_{ikj} = \delta_{ik} + \delta_{kj} \tag{2}$$

The author defines the *best relay path* as the path that has the smallest end-to-end latency among all possible relay paths. The end-to-end latency of the best relay path can be described as follows.

$$\hat{D}^2_{ij} = \min_{k \neq i,j} \left(D^2_{ikj} \right) \tag{3}$$

The author also introduces the *improvement ratio*, which is the ratio of the performance of the best relay path to that of the direct path. The improvement ratio for end-to-end latency is defined as follows.

$$I(D_{ij}^2) = \frac{D_{ij}^1}{\hat{D}_{ij}^2} \tag{4}$$

3.2 Available bandwidth

The overlay routing based on the available bandwidth is suitable for applications that generate a large amount of traffic such as video streaming [24, 25] and file transmission [26]. The available bandwidth of the overlay link between nodes i and j is denoted as β_{ij}. For the available bandwidth of a relay path, the smallest bandwidth between the overlay links that make up the relay path is utilized. Then, we determine the available bandwidth of the direct path between nodes i and j, which is denoted as B_{ij}^1, and the available bandwidth of the relay path via node k, B_{ijk}^2, respectively, as follows.

$$B_{ij}^1 = \beta_{ij} \tag{5}$$

$$B_{ikj}^2 = \min\left(\beta_{ik}, \beta_{kj}\right) \tag{6}$$

The available bandwidth of the best relay path, which has the largest available bandwidth among all possible relay paths, can be described as follows.

$$\hat{B}_{ij}^2 = \max_{k \neq i,j}\left(B_{ikj}^2\right) \tag{7}$$

The improvement ratio for the available bandwidth, which is the ratio of the performance of the best relay path to that of the direct path, is defined as follows.

$$I(B_{ij}^2) = \frac{\hat{B}_{ij}^2}{B_{ij}^1} \tag{8}$$

3.3 Inter-ISP transit cost

Since inter-ISP transit cost is generated by the traffic which traverses inter-ISP transit links, we can reduce the inter-ISP transit cost in overlay routing by decreasing the number of transit links on a path used by overlay networks. Therefore, the author considers the number of transit links on a path chosen by the overlay routing as a metric of the inter-ISP transit cost.

In what follows, τ_{ij} represents the number of transit links of the path between nodes i and j. Then, the number of transit links of the direct path between nodes i and j and that of the relay path via node k are given, respectively, as follows.

$$
\begin{aligned}
T_{ij}^{1} &= \tau_{ij} \qquad (9) \\
T_{ikj}^{2} &= \tau_{ik} + \tau_{kj} \qquad (10)
\end{aligned}
$$

The author proposes the estimation method of the number of transit links in Section 5, since there is no simple end-to-end method to obtain the number of transit links on an end-to-end path.

3.4 Limited overlay routing to reduce inter-ISP transit cost

To reduce the inter-ISP transit cost by overlay routing, the author proposes the following constraint on the overlay routing where the metric explained in Subsection 3.3 is utilized.

$$
T_{ikj} \leq T_{ij} + \alpha \qquad \alpha = 0, 1, \ldots, n \qquad (11)
$$

where α is the upper limit of the increase in the number of transit links utilized by a relay path instead of a direct path. This means that the overlay routing mechanism would choose overlay-level paths which increase the metric value in Subsections 3.1 and 3.2, under the limitation in the increase degree of the number of inter-ISP transit links.

4 Dataset

In this thesis, the author assumes two overlay network environments. One is the overlay network which is constructed of PlanetLab [27] nodes. The other is the overlay network which is constructed of nodes located at Japanese commercial ISPs. To evaluate the overlay routing and the proposed method in each environment, the author needs to know the following properties on the end-to-end path between overlay nodes: end-to-end latency, available bandwidth, router-level path and hop count, and AS-level path and hop count. The author also needs the information of the transit/peering relationships between ISPs to evaluate the transit cost of the overlay routing. In what follows in this section the author explains how to obtain those values for each environment.

4.1 PlanetLab environment

In PlanetLab environment, the author obtained the dataset with the methods described below.

End-to-end latencies: We obtain latency of end-to-end path between PlanetLab nodes from Scalable Sensing Service (S^3) [28]. In S^3, the measurement results for all network paths between PlanetLab nodes, which are summarized every four hours, are available. For end-to-end latency, S^3 uses two types of latencies, one is *measured_latency* which is actual measured values and another is *nv_estimated_latency* which is estimated by Netvigator proposed by Sharma et al. in [29]. Since *measured_latency* is not available about many node pairs, *nv_estimated_latency* is used in this thesis. To avoid the effect of day-to-day fluctuation of the measurement results, the author mainly used the two weeks' datasets obtained from November 12 2008 to November 25 2008 and calculated the median of the datasets. In addition, the datasets on October 25 2006, September 2 2007, and April 8 2009 are also used to evaluate year-on-year changes in the effect of overlay routing in Subsection 6.1.

Available bandwidths: They are obtained with the same way as end-to-end latencies. The measurement result of available bandwidth in S^3 is obtained with two type of tools, pathChirp [30] and Spruce [31]. In this thesis, Spruce data is used since pathChirp data may have overestimated values, which is supposed by the fact that their values are larger than the bottleneck link capacity measured by Pathrate [32] in S^3 in roughly 60% node pairs. In the case of Spruce, the ratio is around 20%.

IP-level paths and router-level hop count: The author conducted `traceroute` commands between all node pairs in PlanetLab. In this thesis, the `traceroute` results obtained on November 12 2008 are used.

AS-level paths and AS-level hop count: The author converts the IP-level path to AS-level path by using the relationships between IP address prefix and AS number which is available at the Route Views Project [33].

Transit/peering information: To obtain the number of transit links on each path, the author utilizes the transit/peering relationship information between ASes which is available at CAIDA [34]. This information is obtained with the method in [35, 36]. However, CAIDA does not provide the relationship information for all links between ASes and there are many IP addresses for routers whose corresponding AS numbers cannot be obtained by the method described above. Therefore, the author applies two additional methods to infer the relationship information. One of the methods is based on the degree of each AS (the number of outgoing links to other ASes). The author first obtains the degree of each AS from CAIDA and derives the ratio at which the relationship is peering for each pair of degrees of ASes. Figure 3 depicts the distribution of the ratio for various pair of ASes' degrees. Then, the unknown relationship information is stochastically determined according the ratio distribution. The other method is based on the property of Boarder Gateway Protocol (BGP). When BGP does not advertise the AS number of a router (i.e., the AS number of a router cannot be obtained with above-described method) on the IP-level path obtained with the above-described method, it may represent that there is no need to be advertised since the router belongs to the same AS where the previous-hop router is located. For this reason, as depicted in Figure 4, when there exists an IP-level path which is constructed of the router of AS X, the router whose AS number is not advertised, and the router of AS Y, then the relationships between each router are estimated respectively as peering and as the relationship between AS X and AS Y.

In this thesis, the number of transit links obtained with above methods is assumed as the "true" number of transit links, and they are utilized for both of the analysis in Section 5 and performance evaluation in Section 6.

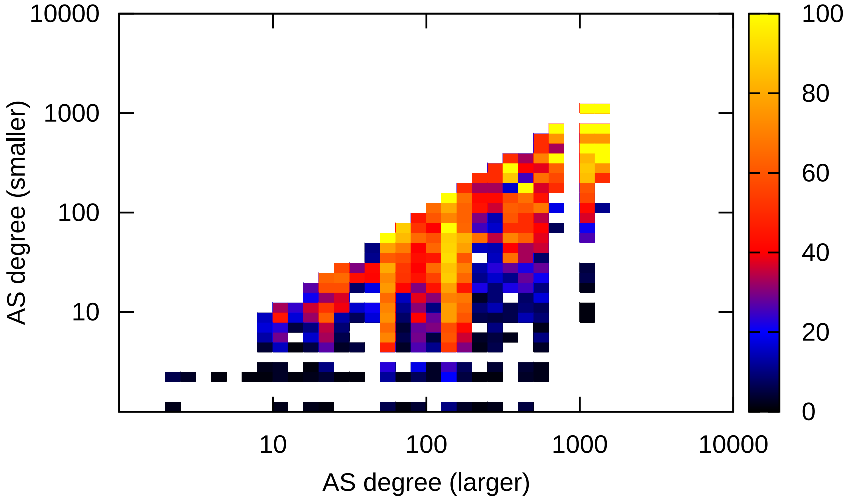


Figure 3: Peering ratio of each pair of ASs' degrees

4.2 Japanese commercial network environment

The dataset on Japanese commercial network environment is obtained from one of our research colleague. The dataset includes the full-mesh `traceroute` results and end-to-end latency measured with `ping` commands, for 18 nodes located at 13 Japanese commercial ISPs, so end-to-end latency and IP-level paths can be obtained from the dataset. Since the dataset does not include the data on available bandwidth, the evaluation on the available bandwidth-based overlay routing are excluded in Section 6. In this thesis, the dataset obtained on March 22 2009 is utilized.

The other data, such as AS-level paths and transit/peering information, are obtained by the identical way in Subsection 4.1.

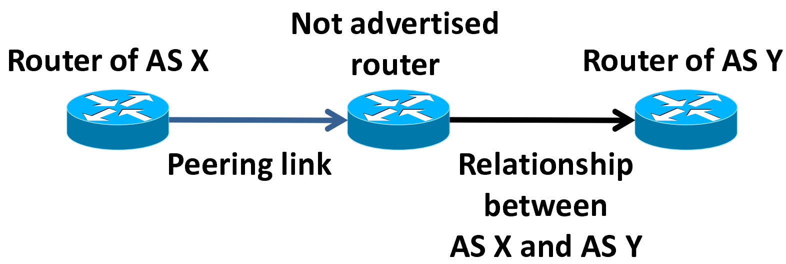


Figure 4: Relationships inferred from property of BGP

5 Estimation of number of transit links on a network path

As described above, the number of transit links on the path between the source and destination overlay nodes is utilized for inter-ISP transit cost. However, the number of transit links cannot be known by overlay nodes because the contract information between ISPs is not disclosed in general. Furthermore, there is no effective method to measure the number of transit links in an end-to-end manner. Indeed, in [35], the relationships between ISPs are inferred by collecting BGP messages from numerous backbone routers. Therefore, the author proposes a method for estimating the number of transit links on a path from other network performance values, which can be measured easily by overlay nodes.

The proposed estimation method first calculates the correlation coefficients between the true number of transit links on a path and three kinds of network performance values to select parameters for multiple regression analysis. After that, the regression equation is derived by using the selected parameters. In what follows in this section, the detailed process of the estimation is described. The author also gives the regression equations for the PlanetLab environment and Japanese commercial network environment and evaluates the estimation accuracy of the regression equation. Besides, the author investigates the difference in the network properties in the PlanetLab environment and Japanese commercial environment.

5.1 Correlation between number of transit links and network performance values

To select parameters for multiple regression analysis, the author first evaluates the correlations between the true number of transit links and network performance values, which are end-to-end latency, available bandwidth, and router-level hop count. For this purpose, *Pearson's correlation coefficient* [37] C in Equation (12) is utilized.

$$C = \frac{\sum (x_{ij} - \bar{x})(y_{ij} - \bar{y})}{\sqrt{\sum (x_{ij} - \bar{x})^2}\sqrt{\sum (y_{ij} - \bar{y})^2} \qquad (12)$$

where x_{ij} is the number of transit links and y_{ij} is each performance value (i.e., router-level hop count, end-to-end latency, and available bandwidth) on the path between nodes i and j. Table 1 lists the correlation coefficients whose values are between the true number of transit links and each performance value in the PlanetLab environment. Unlike the correlation against the router-level hop count and the end-to-end latency, the correlation between the number of transit links and the

Table 1: Correlation coefficients

Router-level hop count	0.420
End-to-end latency	0.300
Available bandwidth	-0.027

available bandwidth is close to zero. From the viewpoints of calculation complexity and accuracy of regression analysis, available bandwidth is excluded from the multiple regression analysis.

For the Japanese commercial environment, there is no data on available bandwidth as mentioned above, so the same parameters (i.e., router-level hop count and end-to-end latency) are selected.

5.2 Regression equation by multiple regression analysis

For the reasons described above, router-level hop count and end-to-end latency are utilized for multiple regression analysis to estimate the number of transit links on a path. The author deploys the linear least squares method to derive the regression equation. S represents the matrix of sum of squares and products of the deviations between performance values and c represents the vector of sum of products of the deviations between the true number of transit links and each performance values. Then, b, which is the partial coefficient vector, is described as follows.

$$b = S^{-1}c \tag{13}$$

We denote the average number of transit links, router-level hop count, and end-to-end latency as $\bar{t}$, $\bar{r}$, and $\bar{d}$, respectively, and b_r and b_d represent partial coefficient values for each parameter. Then b_y, which is the intercept of the regression equation is described as follows.

$$b_y = \bar{t} - \bar{r}b_r - \bar{d}b_d \tag{14}$$

We also denote h_{ij} and δ_{ij} as the router-level hop count and the end-to-end latency (ms) between nodes i and j, respectively. Then we can calculate the regression equation to estimate the number of transit links on a path, denoted as T^e_{ij}, is described as follows.

$$T^e_{ij} = b_r h_{ij} + b_d \delta_{ij} + b_y \tag{15}$$

Table 2: Partial coefficient values of the regression equation

	b_r	b_d	b_y
PlanetLab	0.1353	0.002626	1.216
Japanese commercial network	0.2404	-0.0008878	-1.480

Table 2 shows the partial coefficient values for the PlanetLab environment, using the median of the two weeks' datasets explained in Section 3 and for the Japanese commercial network environment.

5.3 Estimation accuracy of regression equations

The author shows the evaluation results of the estimation accuracy of the regression equation to confirm the effectiveness of the analysis. For this purpose, an estimation error between the true number and the estimated number of transit links on a path is calculated for each overlay node pair. Note that the estimation error considers positive and negative of each value. The true number of transit links on the path between nodes i and j is denoted as T^t_{ij}, and the estimated number of transit links with the regression equation is represented by T^e_{ij}. Then the estimation error of the path between nodes i and j, d_{ij}, is obtained as follows.

$$d_{ij} = T^e_{ij} - T^t_{ij} \tag{16}$$

Figure 5 plots the cumulative distribution of d_{ij} for the paths of all node pairs in PlanetLab environment. For comparison purpose, the results for the single regression analysis using the router-level hop count and the end-to-end latency are plotted, respectively. The figure indicates that the maximum absolute underestimation error of the regression equation in Equation (15) is smaller than four and that the absolute underestimation error is smaller than one for almost 80% of overlay paths. Furthermore, compared to the result obtained using the single regression analysis, the multiple regression equation can give more accurate estimation.

5.4 Network properties affecting regression equations

We can observe the difference between the regression equations on each environment as shown in Table 2. The router-level hop count weighs heavily in the Japanese commercial network environment than the PlanetLab environment and the equations on each environment have approximately

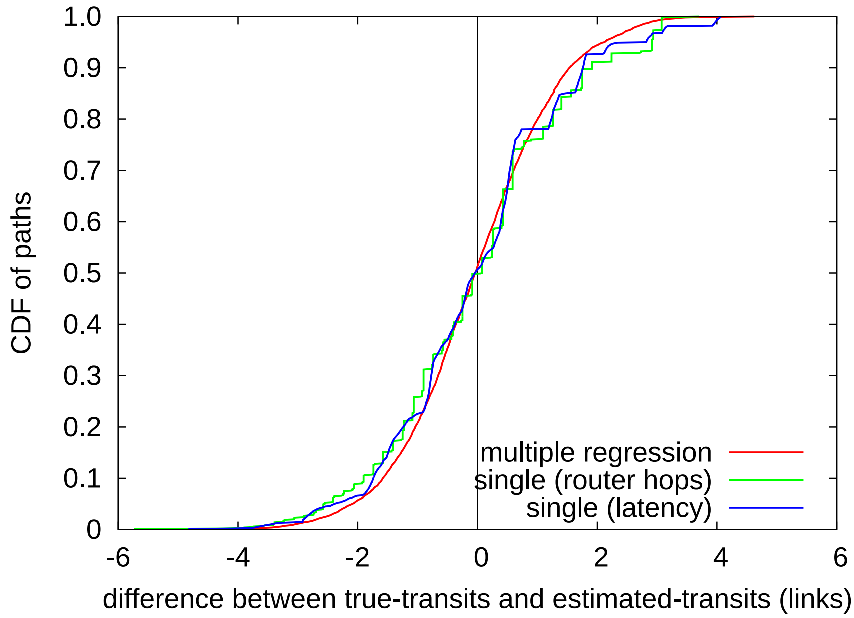


Figure 5: Estimation error distribution of the regression equation

the same intercepts in absolute value but the value in the PlanetLab environment is positive, while that in the Japanese commercial network is negative.

What makes these differences can be found in Figures 6 and 7. Figure 6 shows the distributions of end-to-end latencies of all node pairs' direct path in each environment. Figure 6 tells that the latencies on Japanese commercial network are significantly smaller than that in the PlanetLab environment, so the transit-link estimation by end-to-end latency is difficult in the Japanese commercial network. Therefore, the router-level hop count weighs heavily in the Japanese commercial network environment.

Figure 7 represents the cumulative distribution of the router-level hop count where the path traverses the first transit link for all node pairs' direct paths. The paths which have no transit links are counted at the right endpoint of the x-axis. In the PlanetLab environment, 15% of all paths traverse the first transit links within the beginning first three hops and 47% traverse within five hops. In the Japanese commercial network environment, on the other hand, the values are 0% and 13%, respectively. This means that the paths in the PlanetLab environment have the first transit links within the beginning several hops, so the equation has the positive intercept. On the other hand, since the paths in the Japanese commercial network environment have no transit links in the

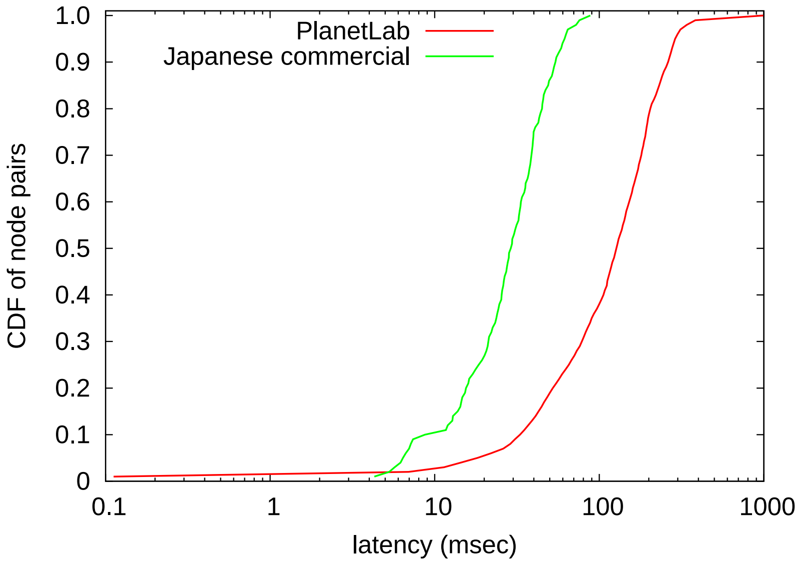


Figure 6: End-to-end latency distribution

beginning part of the path, the equation has the negative intercept.

These differences of network properties may be caused by the differences between the PlanetLab environment and the Japanese commercial network environment. PlanetLab is a global research network and constructed of the nodes which are in universities and enterprises, while the Japanese commercial network is constructed of the nodes located at Japanese commercial ISPs. The proposed method can obtain the regression equations appropriate to each network property.

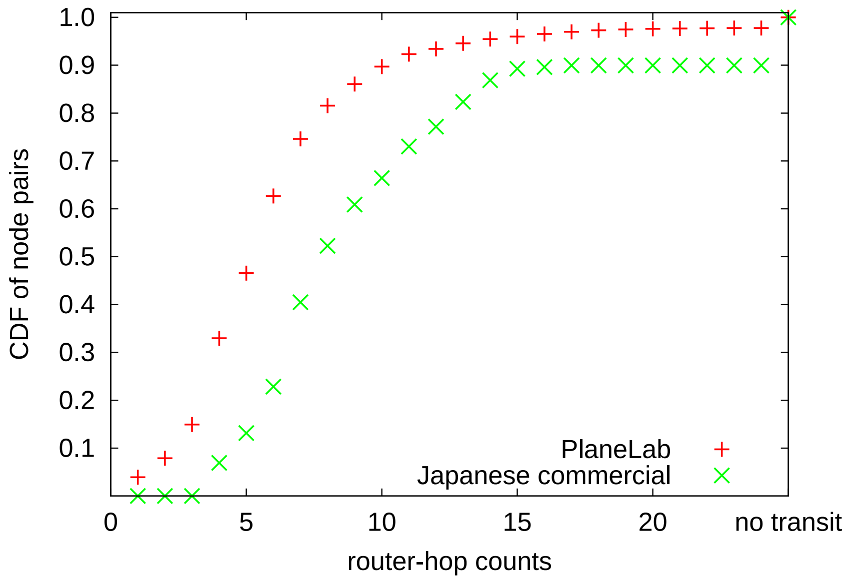


Figure 7: Router-level hop count of the first traversing transit link

6 Numerical Evaluation of overlay routing

In this section, the author first evaluates the performance improvement of the overlay routing without limitation on transit links as a baseline for the discussion. Next, to confirm the effectiveness of the limitation on transit links, the author evaluates the overlay routing with the limitation on the *true* number of transit links. Then, the author shows the results for the performance of the overlay routing with the limitation on the *estimated* number of transit links to confirm the effectiveness of the estimation method in Section 5 and the limited overlay routing proposed in Subsection 3.4.

In the evaluation, the author assumes the PlanetLab environment and the Japanese commercial network environment as mentioned in Section 4. Since there is no data on available bandwidth, the evaluation of available bandwidth-based overlay routing is excluded in the Japanese commercial network environment.

6.1 Evaluation results without transit link limitation

Figure 8 shows the distribution of the relationships between the performance of the direct path and that of the best relay path, which is the 2-hop relay path which has the best performance, for each node pair in the PlanetLab environment. The results with end-to-end latency and available bandwidth utilized as routing metric are shown in Figures 8(a) and 8(b), respectively. In each graph, x-axis represents the performance of the direct path and y-axis represents that of the best relay path corresponding to the direct path. The results in Figure 8 are based on the S^3's measurement results on November 12 2008. These figures show that there is little difference between the direct path and the best relay path in the case of end-to-end latency (Figure 8(a)). This means that end-to-end latency-based overlay routing may improve user-perceived performance only slightly. On the other hand, most node pairs have the best relay path which has larger performance than the direct path in the case of available bandwidth. This result agrees with the results in [16], meaning that available bandwidth-based overlay routing may improve user-perceived performance significantly.

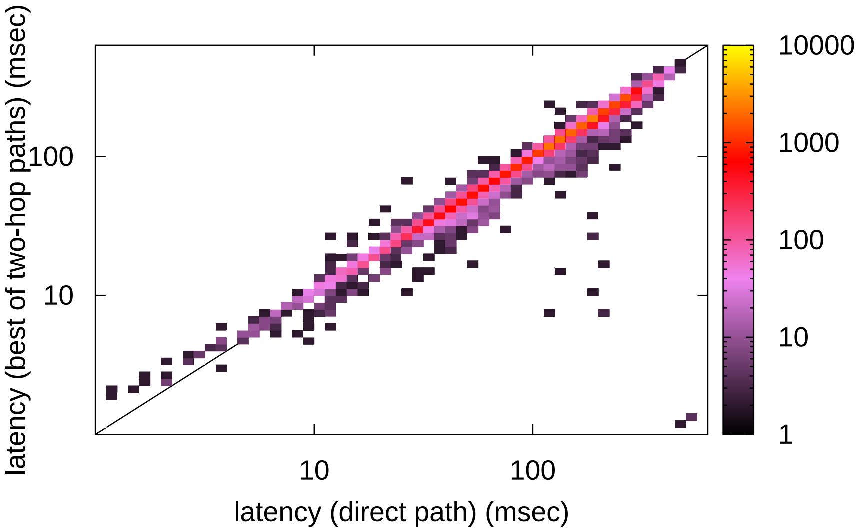
Figure 9 plots the cumulative distribution of the improvement ratio, as defined in Section 3, for all node pairs, when using end-to-end latency (Figure 9(a)) and available bandwidth (Figure 9(b)), respectively. In order to investigate year-on-year changes, the figures include the results based on the datasets obtained on October 25 2006, September 2 2007, November 12 2008, and April 8 2009. In the case of end-to-end latency (Figure 9(a)), the overlay routing shows the best performance on October 25 2006, and the performance decreases on September 2 2007 and November 12 2008. On April 8 2009 the performance is similar to the performance on November 12 2008. The ratios of node pairs that have at least one relay path that is better than the direct path are 67%, 63%, 22%, and 22%, respectively. One possible reason is the decrease in the degree of "distorted" routing configurations caused by commercial inter-ISP relationships. On the other hand, in the case of available bandwidth, although the performances of the overlay routing on November 12 2008 and April 8 2009 are better than the performances on October 25 2006 and September 2 2007, there are significant improvements for all years. The ratio of node pairs that have at least one relay path that is better than the direct path is larger than 95% for all years.

Figure 10 shows the result in the same manner as Figure 9 for the Japanese commercial network environment. Since we do not have the data about available bandwidth for this environment, the evaluation is only on the end-to-end latency-based overlay routing. The ratio of node pairs that have at least one relay path that is better than the direct path is 15%. The tendency of the results is similar to that with the datasets on November 12 2008 and April 8 2009 in the PlanetLab environment in Figure 9.

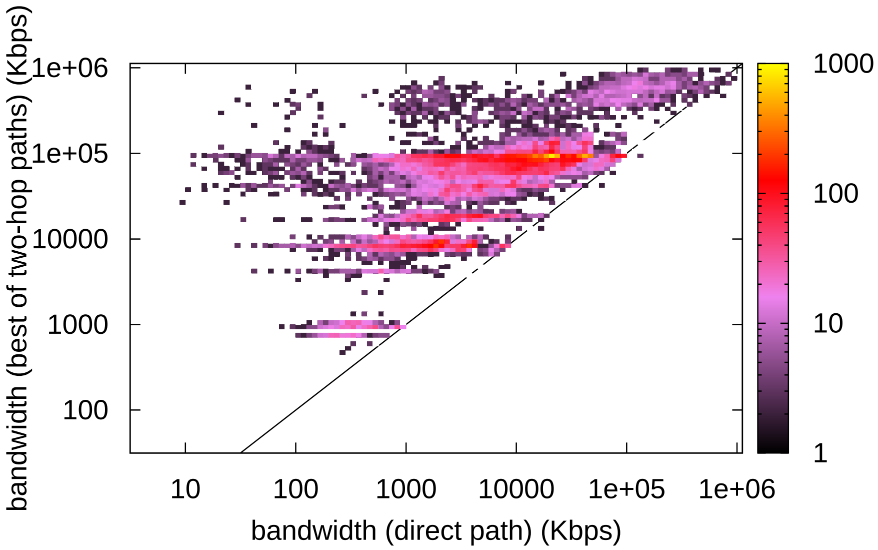
6.2 Evaluation results with limitation on true number of transit links

Next, the author shows the results for the case with the limitation on the true number of transit links which is explained in Section 4. The detailed algorithm of limited overlay routing can be found in Section 3.

Figure 11 exhibits the cumulative distribution of the improvement ratio of all paths when limiting the increase in the number of transit links, where α is the upper limit of the increase in the number of transit links, described in Subsection 3.4. The results in Figure 11 are based on the dataset which is the median of two weeks' datasets obtained from November 12 2008 to November 25 2008. Note that when α is small, we cannot find any relay paths that satisfy the limitation for some node pairs. Figure 11 also indicates that, for the both case of using end-to-end latency and available bandwidth as routing metrics, as α increases, the performance approaches that for the case without limitation, and when α is greater than or equal to three, the performances become approximately equal. From these results, we conclude that the overlay routing with the limitation on the number of transit links can provide the performance similar to the case without the limitation, when the limitation on the increase in the number of transit links is greater than or

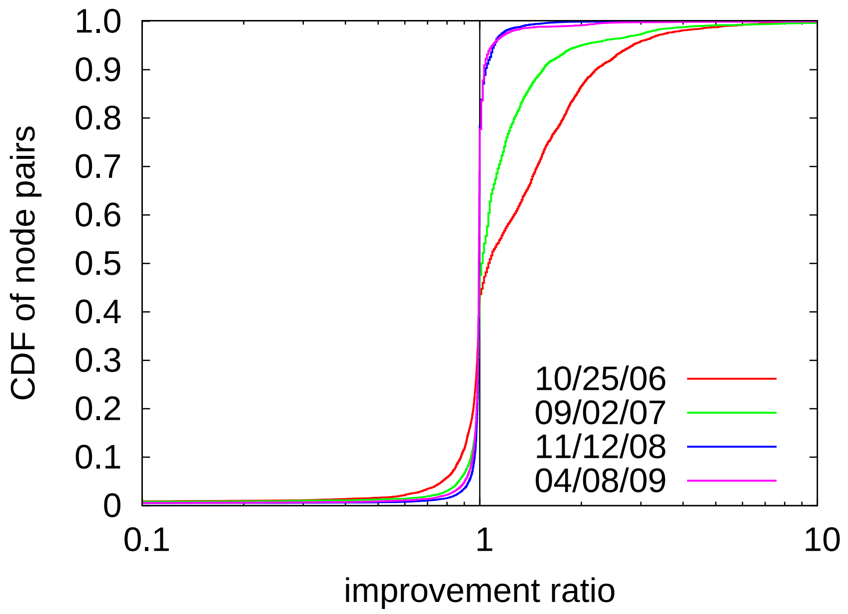


(a) end-to-end latency

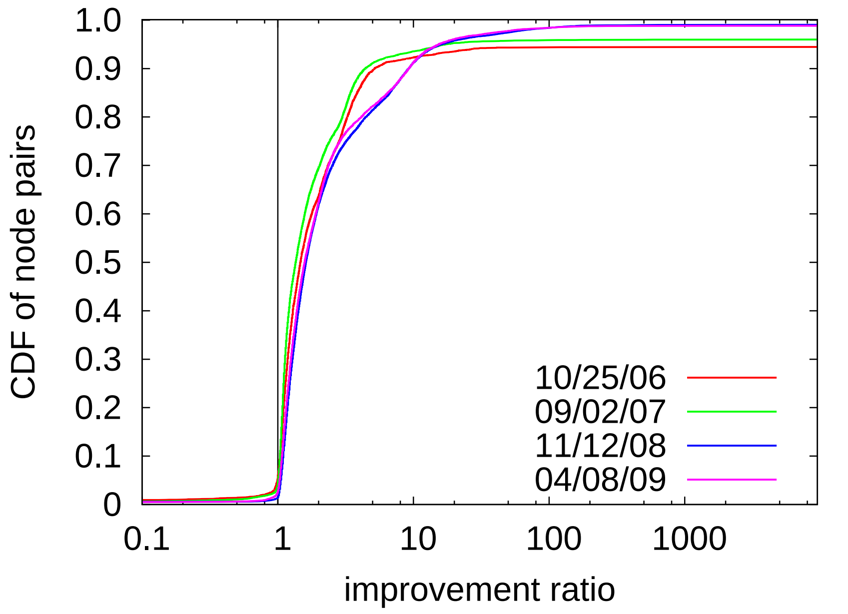


(b) available bandwidth

Figure 8: Comparison between the direct path and the best relay path



(a) end-to-end latency



(b) available bandwidth

Figure 9: Year-on-year changes in improvement ratio distribution in PlanetLab environment

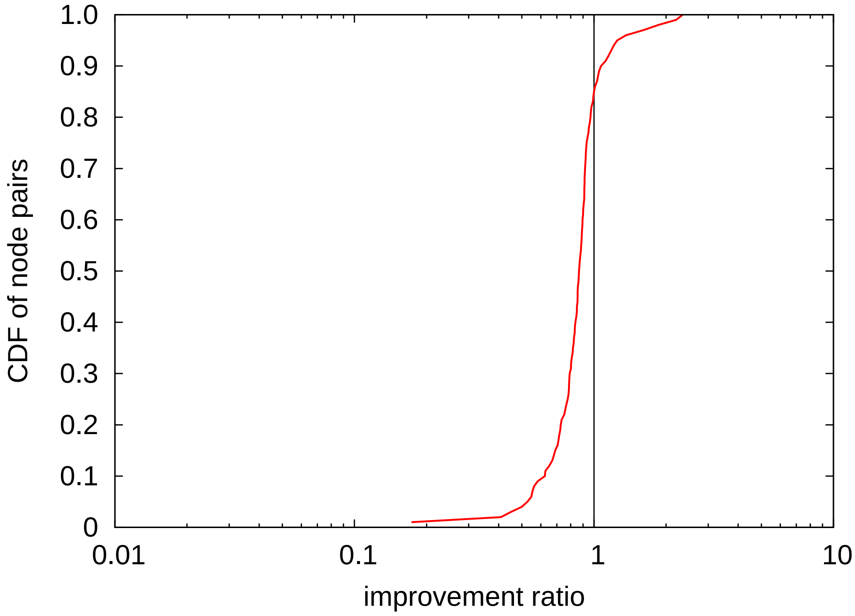


Figure 10: Improvement ratio distribution in Japanese commercial network environment

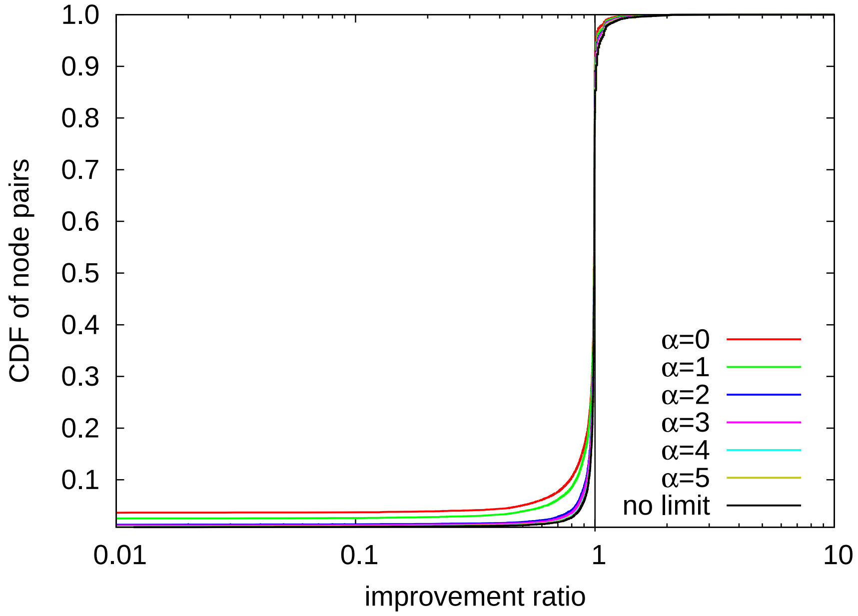
equal to three.

Figure 12 shows the results in the same manner as Figure 11 for the Japanese commercial network environment. Although the tendency of the results is similar to that in the PlanetLab environment except for one difference. That is, when α is greater than or equal to one (three for PlanetLab environment), the performance improvement is approximately equal to that without the limitation. It is due to the difference in network property between both environments, which is estimated in Table 2. That is, the number of transit links of paths in the Japanese commercial network environment is smaller than that in the PlanetLab environment.

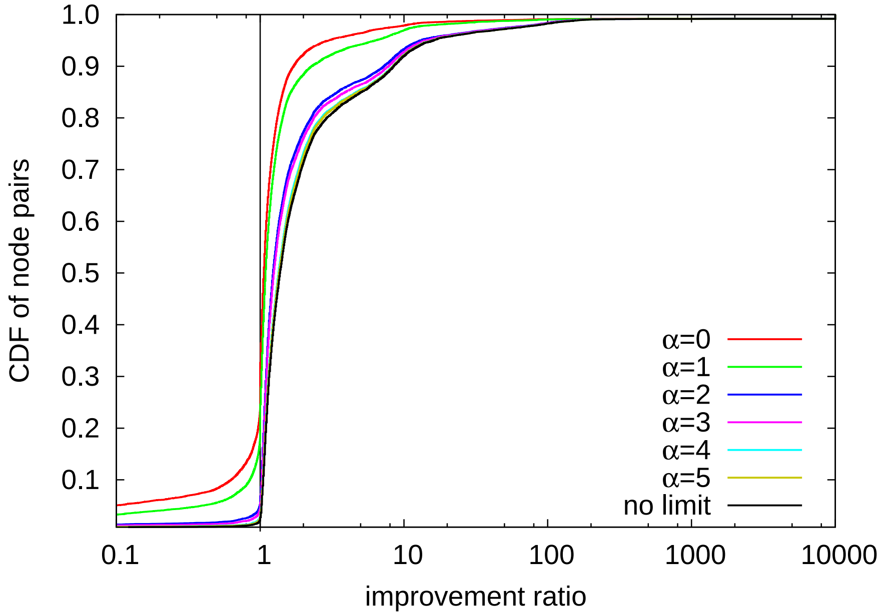
6.3 Evaluation results with limitation on estimated number of transit links

Finally, the author shows the evaluation results of the performance improvement of the overlay routing with the limitation on transit links estimated by the proposed method to confirm the effectiveness of the proposed method. The datasets used in this subsection is identical to that in Subsection 6.2.

Figure 13 plots the results in the same manner as Figure 11 with the estimated number of transit links instead of the true number of transit links. This figure tells that when α is smaller than three,



(a) end-to-end latency



(b) available bandwidth

Figure 11: Improvement ratio distribution with the limitation on the true number of transit links (PlanetLab environment)

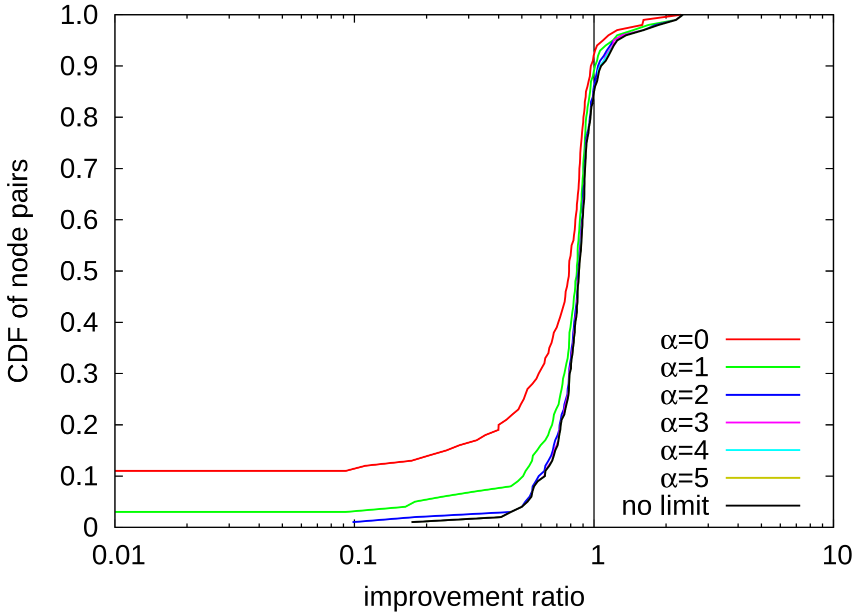


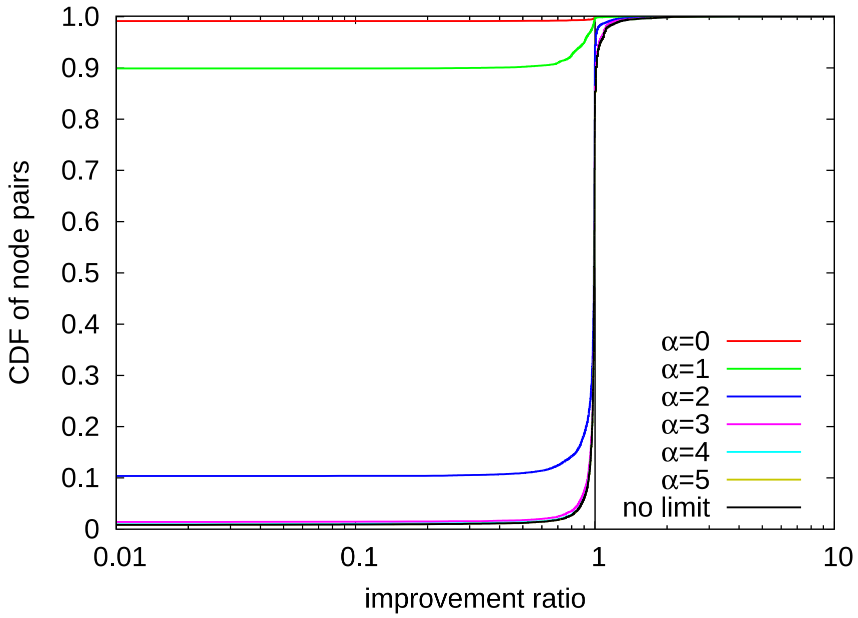
Figure 12: Improvement ratio distribution with the limitation on the true number of transit links (Japanese commercial network environment)

there are many node pairs whose improvement ratio is zero in the figure and the portion increases significantly compared with Figure 11, since a significant portion of the node pairs cannot find any relay paths which satisfy the limitation. This is because of the estimation error described in Subsection 5.3. On the other hand, when α is greater than or equal to three, the overlay routing performance is approximately the same as in the case with the true number of transit links (Figure 11) and the case without the limitation (Figure 9).

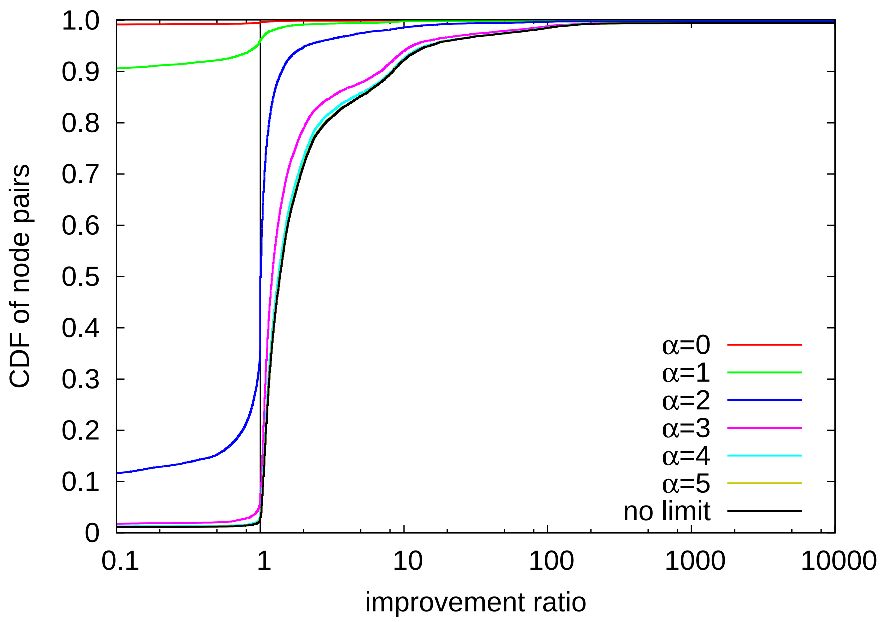
Figure 14 shows the results in the same manner as Figure 13 for the Japanese commercial network environment. The tendency of the results is similar to that in the PlanetLab environment. When α is greater than or equal to one, the overlay routing performance is approximately the same as the case with the true number of transit links (Figure 12) and the case without the limitation (Figure 10). The results reveal the same advantage as for the PlanetLab environment.

6.4 Effect of geographic distribution of overlay nodes

One of possible discussions on the above results is that the results on the performance of the overlay routing may be biased by the skewed distribution of the locations of PlanetLab nodes. Indeed,



(a) end-to-end latency



(b) available bandwidth

Figure 13: Improvement ratio distribution with the limitation on the estimated number of transit links (PlanetLab environment)

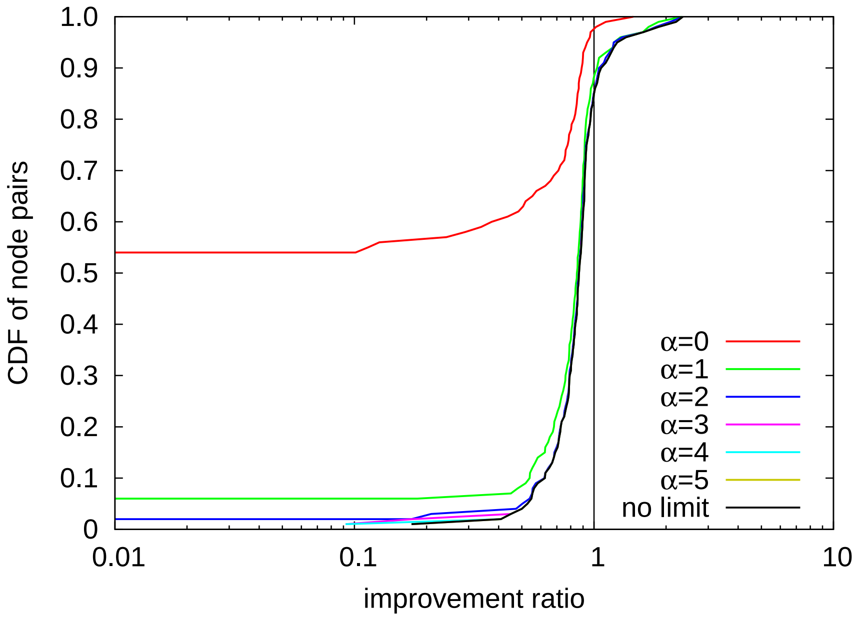
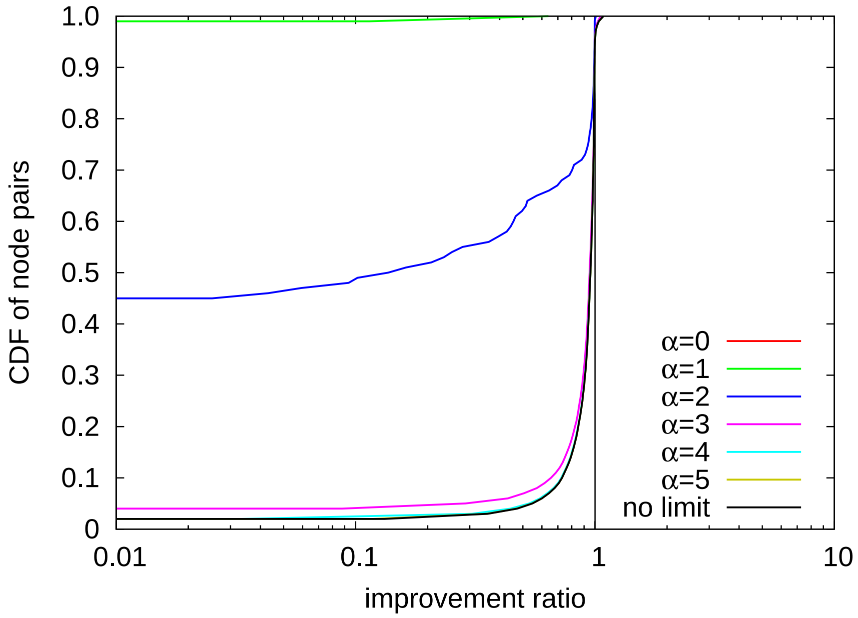


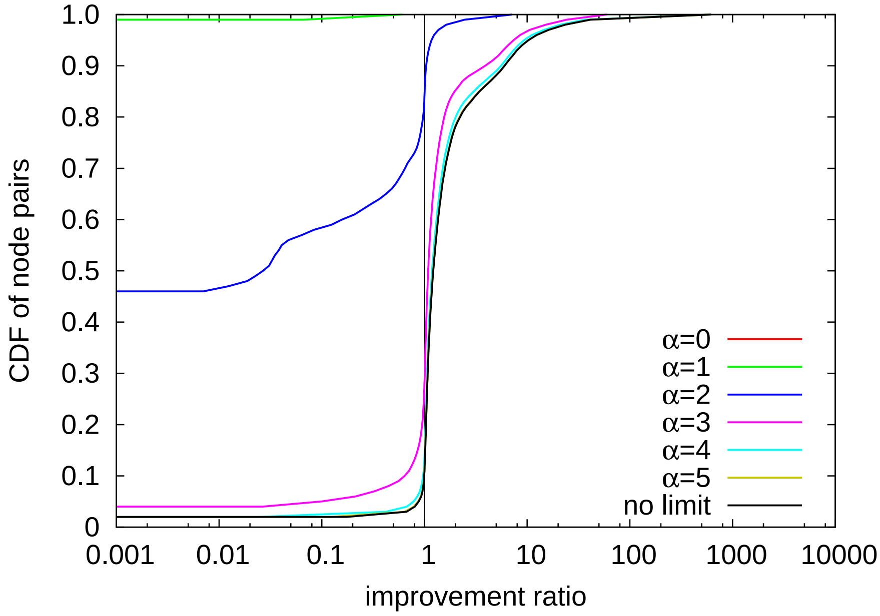
Figure 14: Improvement ratio distribution with the limitation on the estimated number of transit links (Japanese commercial network environment)

most of PlanetLab nodes are located at North America and Europe. To confirm the effect of the geographic distribution of overlay nodes, the author finally investigated the performance of the overlay routing where the overlay node distribution is tuned to the Internet host distribution. For this purpose, the number of ASes in each Regional Internet Registry (RIR) in the current Internet environment is utilized. Table 3 summarizes the distribution of the number of ASes obtained from [38] and the number of utilized PlanetLab nodes in each region, which is determined proportionally to the number of ASes. We randomly selected the PlanetLab nodes from each region and evaluated the performance of the overlay routing which is operated by the selected nodes.

Figure 15 plots the results of distribution of the improvement ratio in the overlay routing with selected nodes. By comparing Figures 13 and 15, the tendency of the results is almost identical that with all nodes in the PlanetLab environment, especially when α is larger than or equal to three. From these results, the author confirmed that the proposed method is effective for the overlay routing not only in the PlanetLab environment which is mainly constructed of nodes in North America and Europe, but also in the normal network environment.



(a) end-to-end latency



(b) available bandwidth

Figure 15: Improvement ratio distribution with the limitation on the estimated number of transit links (selected nodes)

Table 3: Number of ASes in each RIR and number of nodes for evaluation

RIR (region name)	number of ASes	number of nodes
ARIN (North America)	24422	50
RIPE NCC (Europe)	21065	43
APNIC (Asia)	5782	12
LACNIC (South America)	2815	6

7 Conclusion

In this thesis, the author proposed a method to reduce inter-ISP transit cost caused by overlay routing. The proposed method estimates the number of transit links on a path using multiple regression analysis of end-to-end network performance values which can be measured easily by overlay nodes and limits the increase of transit links on overlay-routed paths to reduce inter-ISP transit cost. Through the extensive evaluation using measurement results in the actual network environment, the author confirmed the effectiveness of the proposed method. The absolute underestimation error of the proposed estimation method is less than one in almost 80% of overlay paths and the overlay routing with the proposed method can achieve almost the same degree of performance improvement as that without the transit link limitation, while significantly decreasing the number of transit links traversed by overlay-routed paths. The results revealed the advantage of the proposed method whereby we can control the number of transit links in overlay routing using measurable network performance values, while preserving the performance improvement by the overlay routing.

In the future, the author intends to consider different mechanisms to decrease inter-ISP transit cost by explicit cooperation between ISPs and overlay network applications, such as [39, 40]. These mechanisms can optimize routing of intra-ISP traffic in each ISP, but inter-ISP traffic is not considered. Therefore, after investigating properties of these mechanisms, the author plans to build up a mechanism considering routing of both intra-ISP and inter-ISP traffic by combining the proposed method in this thesis and these mechanisms.

Acknowledgement

I would express my thanks to Professor Hirotaka Nakano of Osaka University. His comments polished my study.

I am deeply grateful to Associate Professor Go Hasegawa of Osaka University. All my works have been supported greatly by him. Without his support, my works would not be achieved. I want to appreciate sincerely for his ardent leading and precious advices.

I am also deeply grateful to Professor Masayuki Murata of Osaka University. His pointed and worthful advices gave me the direction of my study and brushed my works up.

My comfortable research life is offered by Assistant Professor Yoshiaki Taniguchi of Osaka University. I want to thank for his support.

My life in master's course has been worthwhile days because of the meaningful guidance by all professors of Graduate School of Information Science and Technology, Osaka University. I express my sincere appreciation to them.

I also give my thanks to Masahumi Hashimoto, my colleague in Nakano laboratory. His comments and advices are very useful for my works.

Finally, I express my special thanks to all my friends and my family. They have encouraged me at all times.

References

[1] J. Shamsi and M. Brockmeyer, "QoSMap: Achieving quality and resilience through overlay construction," in *Proceedings of ICIW 2009*, pp. 58–67, May 2009.

[2] Z. Li and P. Mohapatra, "QRON: QoS-aware routing in overlay networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, pp. 29–40, Jan. 2004.

[3] S. Banerjee, C. Kommareddy, K. Kar, B. Bhattacharjee, and S. Khuller, "Construction of an efficient overlay multicast infrastructure for real-time applications," in *Proceedings of INFOCOM 2003*, Apr. 2003.

[4] S. Shi and J. Turner, "Routing in overlay multicast networks," in *Proceedings of INFOCOM 2002*, June 2002.

[5] A. Nakao, L. Peterson, and A. Bavier, "A routing underlay for overlay networks," in *Proceedings of SIGCOMM 2003*, pp. 11–18, Aug. 2003.

[6] R. Kawahara, E. K. Lua, M. Uchida, S. Kamei, and H. Yoshino, "On the quality of triangle inequality violation aware routing overlay architecture," in *Proceedings of INFOCOM 2009*, pp. 2761–2765, Apr. 2009.

[7] D. G. Andersen, A. C. Snoeren, and H. Balakrishnan, "Best-path vs. multi-path overlay routing," in *Proceedings of IMC 2003*, Oct. 2003.

[8] Z. Li and P. Mohapatra, "The impact of topology on overlay routing service," in *Proceedings of INFOCOM 2004*, vol. 1, pp. 408–418, Mar. 2004.

[9] C. Lumezanu, R. Baden, D. Levin, N. Spring, and B. Bhattacharjee, "Symbiotic relationships in Internet routing overlays," in *Proceedings of NSDI 2009*, pp. 467–480, Apr. 2009.

[10] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris, "Resilient overlay networks," in *Proceedings of SOSP 2001*, Oct. 2001.

[11] C. L. T. Man, G. Hasegawa, and M. Murata, "Monitoring overlay path bandwidth using an inline measurement technique," *IARIA International Journal on Advances in Systems and Measurements*, vol. 1, no. 1, pp. 50–60, 2008.

[12] G. Hasegawa, M. Kobayashi, M. Murata, and T. Murase, “Free-riding traffic problem in routing overlay network,” in *Proceedings of ICON 2007*, Nov. 2007.

[13] M. Uchida, S.Kamei, and R. Kawahara, “Performance evaluation of QoS-aware routing in overlay network,” in *Proceedings of ICOIN 2006*, Jan. 2006.

[14] S. Kamei, “Applicability of overlay routing in Japan using inter-domain measurement data,” *Overlay Network Workshop*, Dec. 2006.

[15] Y. Zhu, C. Dovrolis, and M. Ammar, “Dynamic overlay routing based on available bandwidth estimation: A simulation study,” *Computer Networks Journal*, vol. 50, pp. 739–876, Apr. 2006.

[16] G. Hasegawa, Y. Hiraoka, and M. Murata, “Effectiveness of overlay routing based on delay and bandwidth information,” *IEICE Transactions on Communications*, vol. E92-B, pp. 1222–1232, Apr. 2009.

[17] P. Rodriguez, S.-M. Tan, and C. Gkantsidis, “On the feasibility of commercial, legal P2P content distribution,” *SIGCOMM Computer Communication Review*, vol. 36, no. 1, pp. 75–78, 2006.

[18] S. Seetharaman and M. Ammar, “Characterizing and mitigating inter-domain policy violations in overlay routes,” in *Proceedings of ICNP 2006*, pp. 259–268, Nov. 2006.

[19] T. Karagiannis, P. Rodriguez, and K. Papagiannaki, “Should Internet service providers fear peer-assisted content distribution?,” in *Proceedings of IMC 2005*, pp. 6–6, Oct. 2005.

[20] B. Y. Zhao, A. D. Joseph, and J. Kubiatowicz, “Locality aware mechanisms for large-scale networks,” in *Proceedings of Workshop on FuDiCo 2002*, June 2002.

[21] S. Le-Blond, A. Legout, and W. Dabbous, “Pushing BitTorrent locality to the limit,” *Computing Research Repository*, vol. abs/0812.0581, Dec. 2008.

[22] “Skype.” available at `http://www.skype.com/`.

[23] “RealVNC.” available at `http://www.realvnc.com/`.

[24] “YouTube.” available at `http://www.youtube.com/`.

[25] "Joost." available at `http://www.joost.com/`.

[26] "BitTorrent." available at `http://www.bittorrent.com/`.

[27] PlanetLab Web Page. available at `http://www.planet-lab.org/`.

[28] Hewlett-Packard Laboratories, "Scalable Sensing Service." available at `http://networking.hpl.hp.com/s-cube/`.

[29] P. Sharma, Z. Xu, S. Banerjee, and S.-J. Lee, "Estimating network proximity and latency," *SIGCOMM Computer Communication Review*, vol. 36, no. 3, pp. 39–50, 2006.

[30] V. J. Ribeiro, R. H. Riedi, R. G. Baraniuk, J. Navratil, and L. Cottrell, "pathChirp: Efficient available bandwidth estimation for network paths," in *Proceedings of Passive and Active Measurement Workshop*, Apr. 2003.

[31] J. Strauss, D. Katabi, and F. Kaashoek, "A measurement study of available bandwidth estimation tools," in *Proceedings of IMC 2003*, pp. 39–44, Oct. 2003.

[32] C. Dovrolis, M. Jain, and R. Prasad, "BW-meter: measurement tools for the capacity and load of Internet paths." available at `http://www.cc.gatech.edu/fac/Constantinos.Dovrolis/bw-est/`.

[33] University of Oregon, "Route Views Project." available at `http://www.routeviews.org/`.

[34] University of Callfornia, "CAIDA." available at `http://www.caida.org/home/`.

[35] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, k. claffy, and G. Riley, "As relationships: Inference and validation," *SIGCOMM Computer Communication Review*, vol. 37, no. 1, pp. 29–40, 2007.

[36] L. Gao, "On inferring autonomous system relationships in the Internet," *IEEE/ACM Transscations on Networking*, vol. 9, no. 6, pp. 733–745, 2001.

[37] K. Pearson and F. Galton, "Contributions to the mathematical theory of evolution. note on reproductive selection," *Journal of the Royal Statistical Society*, vol. 59, pp. 398–402, June 1896.

[38] IANA, “IANA AS Numbers assignment data page.” available at http://www.iana.org/assignments/as-numbers/.

[39] V. Aggarwal, A. Feldmann, and C. Scheideler, “Can ISPs and P2P users cooperate for improved performance?,” *SIGCOMM Computer Communication Review*, vol. 37, no. 3, pp. 29–40, 2007.

[40] H. Xie, Y. R. Yang, A. Krishnamurthy, Y. G. Liu, and A. Silberschatz, “P4P: Provider portal for applications,” *SIGCOMM Computer Communication Review*, vol. 38, no. 4, pp. 351–362, 2008.