An Enhanced Scalable Proximity Model

Mohammad Malli, Chadi Barakat, Walid Dabbous

An Enhanced Scalable Proximity Model

Mohammad Malli, Chadi Barakat, Walid Dabbous

Projet Planète, INRIA-Sophia Antipolis, France
E-mail:{mmalli, cbarakat, dabbous}@sophia.inria.fr

Abstract. We introduce in this paper the notion of application-level proximity. This proximity is a function of network parameters that decide on the application performance, mainly the delay and the available bandwidth. Most of existing protocols rely on the delay proximity (e.g., the delay closest peer is the best peer to contact). We motivate the need for this new notion by showing that the proximity in the delay space does not automatically lead to a proximity in the bandwidth space. Then, we explain how a landmark approach, designed originally to estimate the delay proximity among peers in a scalable manner, can be extended to account for the available bandwidth. Our solution estimates the bandwidth among peers using the bandwidth of the indirect paths that join them via a small set of landmarks. We evaluate the solution and analyze the impact of the landmarks' locations on the accuracy of the estimation. We obtain satisfactory results when the delay of more than one indirect path is close to that of the direct path. Better results are obtained when more than one landmark are located near one of the path end points. Finally, we show that the proximity determined using our bandwidth estimation model provides much better quality than that obtained using the delay proximity for large file transfer applications. The whole study is based on real measurements conducted over the Planetlab plateform.

1 Introduction

In Peer-to-Peer and overlay networks, the performance perceived by users can be optimized at the application level by identifying the best peer to contact or to take as neighbor. This requires to define a proximity function that evaluates how much two peers are close applicatively to each other.

Different functions are introduced in the literature to characterize the proximity of peers, but most of them [3, 2, 11, 10] are based on simple metrics such as the delay, the number of hops and the geographical location. We believe that these metrics are not enough to characterize the proximity given the heterogeneity of the Internet in terms of path characteristics and access link speed, and the diversity of application requirements. Some applications (e.g., transfer of large files and video streaming) are sensitive to other network parameters such as the bandwidth. Therefore, the proximity should be defined at the application level taking into consideration the network metrics that decide on the application performance. We propose to do that using a utility function that models the quality perceived by peers at the application level. A peer is closer than another one to some reference peer if it provides a better utility function, even if the path connecting it to the reference peer is longer.

Using extensive measurements over the Planetlab overlay network [9], we motivate our work by studying how much a proximity-based ranking of peers using the delay deviates from that using the end-to-end available bandwidth[1]. Our observation is that the delay proximity is not always a good predictor of quality and that the available bandwidth has to be considered as well. Particulary, the best peer to contact is not always the delay closest one. The knowledge of the available bandwidth between peers helps in improving the performance of applications by allowing to define better proximity models. The challenge is to infer the end-to-end available bandwidth in an easy scalable way as it is done for estimating the delay [10–12, 14].

To this end, we explain how a landmark approach, designed originally to estimate the delay proximity among peers in a scalable manner, can be extended to account for the bandwidth. Our solution consists in estimating the bandwidth among peers using the bandwidth of the indirect paths that join them via a small set of landmarks. Again, using Planetlab measurements, we evaluate our solution and analyze the impact of the landmarks' locations on the accuracy of the estimation. We obtain satisfactory results when the delay of more than one indirect path is close to that of the direct path. Better results are obtained when more than one landmark are located near one of the extremities of the path to characterize.

Finally, we compare the delay proximity and application-level proximity from application standpoint. A typical file transfer application is considered to evaluate the degradation of the performance perceived by peers when they choose their neighbors based on these two distinguished proximity notions. We observe that the application-level proximity, which is determined using the landmark-based bandwidth estimation, provides a better quality compared to that obtained when using the delay alone.

The paper is structured as follows. Next we present our measurement setup. Section 3 motivates the need to consider the bandwidth when characterizing the proximity. The scalable model for bandwidth inference is introduced in Section 4. The impact of landmarks' locations on bandwidth estimation accuracy is evaluated in Section 5. Section 6 illustrates the difference in performance between the delay based proximity and the application based one for a typical file transfer application. The paper is concluded in Section 7.

2 Measurement setup

Our experiment consists of real measurements run in February 2005 over the Planetlab platform [9]. We do not claim that this platform is representative of all networks, but we believe that it is the best evaluation testbed available that

[1] The available bandwidth represents the remaining bandwidth left on a path between two nodes. It is determined by the residual bandwidth at the bottleneck link. In the rest of the paper, the term bandwidth is used to denote the available bandwidth.

satisfies the large scale requirement of our study. Moreover, this platform has proved its capacity to be appropriate for measurements [13].

In the rest of paper, we call a Planetlab node a peer. We measure the end-to-end network parameters of the paths connecting peers using the *Abing* tool [7]. This tool is based on the packet pair dispersion technique [6]. It consists of sending a total number of 20 probe packet-pairs between the two sides of the measured path. It has the advantage of short measurement time on the order of the second, a rich set of results (e.g, bandwidth in both directions), and a good functioning over Planetlab. The measurement accuracy provided by Abing on Planetlab is quite reasonable compared to other measurement tools [8].

3 Motivation and Methodology

Different definitions were studied in the literature for characterizing the proximity among peers, and hence for selecting the appropriate peer to contact. These definitions can be classified into two main approaches: static and dynamic. The difference between them lies in the metrics they consider. Static approaches [3, 4] use metrics that change rarely over time as the number of hops, the domain name and the geographical location. Dynamic approaches [1, 2, 11, 10, 14] are based on the measurement of variable network metrics. They mainly focus on the delay and consider it as a measure of closeness among peers; the appropriate peer to contact is often taken as the closest one in the delay space. The focus on the delay is because of its low measurement cost (i.e., measurement time, amount of probing bytes) and the possibility to be estimated in a scalable manner (e.g., using the landmark approach [10, 11, 14, 12]). However, its usage hides the implicit assumption that the path with the closest peer (in terms of delay) has the maximum (or relatively large) bandwidth.

While we believe that the delay can be an appropriate measure of proximity for some applications (e.g., non greedy delay sensitive applications or those seeking for geographical proximity), it is not clear if it is the right measure to consider for other applications whose quality is a function of diverse network parameters. Greedy applications and multimedia ones are typical candidates for a more enhanced definition of proximity. To clarify this point, we check particularly whether (i) the delay and the available bandwidth are correlated with each other, and (ii) how much a proximity-based ranking of peers using the delay deviates from that using the bandwidth.

We take 127 Planetlab nodes spread over the Internet and covering America, Europe, and Asia. Forward and reverse paths between each pair of peers are considered, which leads to 16002 measured paths. For each unidirectional path between two peers, we measure the round-trip time RTT, and the available bandwidth ABw.

For a peer p, we denote by p_d the delay closest peer and by p_a the best peer in terms of available bandwidth. We also denote by $d(x, y)$ (respectively $ABw(x, y)$) the delay (respectively the bandwidth) on the path leading from peer x to peer y. Figure 1 shows the complementary cumulative distribution function (CCDF)

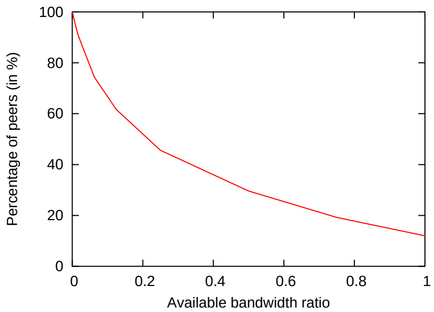

Fig. 1. CCDF of ABw on the delay shortest paths

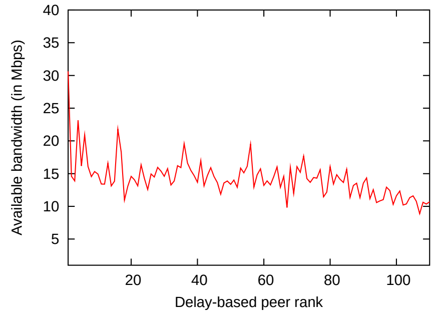

Fig. 2. Variation of ABw with the delay-based rank

of the ratio $ABw(p, p_d)/ABw(p, p_a)$. This illustrates how far is the available bandwidth on the delay shortest path from the maximum available bandwidth. The CCDF is plotted over the 127 peers.

We can see that only (i) 12% of peers have the maximum ABw on their path with the nearest peer, (ii) 19.2% have more than 75% of the maximum ABw, and (iii) 45.6% have more than 25% of the maximum ABw. Thus, the delay is far from being the proximity metric to use to detect the peer with the maximum available bandwidth.

In our setting, the delay and bandwidth are lightly negatively correlated with a coefficient equal to -0.096. This can be observed in Figure 2 where we plot the average available bandwidth for peers of rank r in the delay space, r varying from 1 to 126. We notice that looking at farther and farther peers in the delay space does not lead to an important decrease in the available bandwidth, and so there is a high chance of having the optimal peer from bandwidth point of view located far away (in the delay space) from the peer requesting the service. This low correlation motivates the need for an enhanced model of proximity that is based on the knowledge of the available bandwidth among peers in addition to the delay. Solutions exist in the literature for distributed scalable delay estimation [10–12, 14]. The challenge is how to estimate the bandwidth in an efficient scalable way.

4 Scalable end-to-end bandwidth inference

The end-to-end delay can be estimated easily and scalably between peers using a landmark approach [10–12, 14]. For example, one can calculate coordinates for peers and infer the delay as the Euclidean distance separating them. Peers' coordinates are deduced from delay measurements to a small number N of landmarks $L\{L_1, ..., L_N\}$. We wonder whether it is feasible to use such distributed solution for bandwidth estimation. This requires that each peer also determines its bandwidth vector by measuring the direct and reverse bandwidth on its path with each landmark. If we arrive to design such estimation model, we will be able to infer the bandwidth between peers in a manner which is (i) scalable since the system overhead will be linear with the number of peers in the system, and (ii) easy to implement since peers will not need to know and probe each other; any node can estimate the bandwidth between any two peers based on their bandwidth vectors.

For a couple of peers, we denote by (i) *direct path* the network path that joins them directly using IP routing, and by (ii) *indirect path* the path that joins the two peers by passing by a landmark node[2].

Our idea is to estimate the bandwidth on the direct path using those on the indirect paths. The indirect paths that most probably have common links with the direct path, are assigned more weight in the estimation function. We consider different estimation functions and we study the impact of the landmarks' locations on the accuracy of the estimations.

For a direct path joining two peers, we estimate its end-to-end bandwidth using the following class of linear functions:

$$EB = \sum_{i=1}^{N} P_i \cdot BB_i, \tag{1}$$

where BB_i is the bandwidth of the indirect path that passes by the landmark L_i, and P_i is a normalized weight (i.e., $\sum_{i=1}^{N} P_i = 1$) assigned to this indirect path according to the location of its landmark with respect to the two peers. By varying the weight P_i, we are able to cover different policies for bandwidth estimation ranging from the one that gives the same priority to all landmarks to the one that privileges the landmark that we deem the most suitable for the direct path bandwidth inference.

5 Impact of landmarks' locations

In [14], the authors observe that 8 to 12 landmarks should suffice for a good delay estimation at the scale of the Internet. We consider the same number of landmarks for bandwidth estimation. Therefore, we take 8 Planetlab nodes selected from different European countries as landmarks. We also take 14 Planetlab nodes completely distributed in Europe as peers. Each of these peers measures the RTT and the direct and reverse ABw to 34 Planetlab nodes distributed worldwide. This leads to 476 measured paths. Then, we infer the bandwidth of these paths using Equation (1) and we compare the estimations with the measured values. Furthermore, we study the correlation between the estimation accuracy and the landmarks' locations.

Our landmark nodes are chosen with the main concern to have a high bandwidth connectivity to the Internet. This is an important requirement since we want to avoid having the bottleneck, of an indirect path, decided by the region around the landmark. We want it to be rather decided by the regions around

[2] N indirect paths (N is the number of landmarks) are assigned to each direct path.

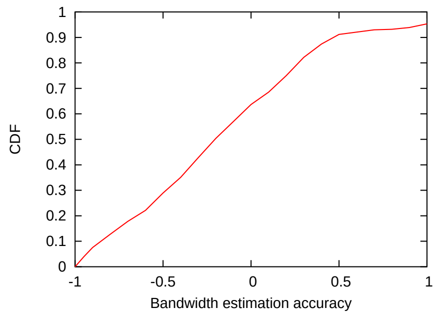

Fig. 3. Bandwidth estimation accuracy

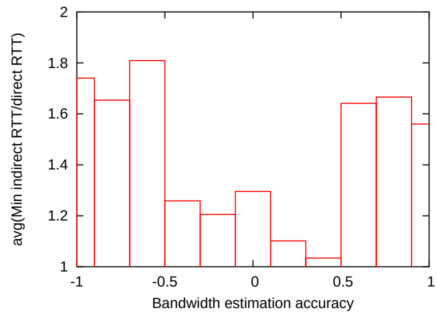

Fig. 4. Bandwidth estimation accuracy

the peers. In fact, it is more probable that the latter regions are common with the direct path compared to that around the landmark.

We consider different forms of the weights P_i, and subsequently of the end-to-end bandwidth estimation function. By doing that, we are able to study the correlation between the estimation accuracy and the locations of the landmarks. We divide the study into two main parts: (i) the estimation function depends on the delay closeness between the direct path and the indirect paths, (ii) the estimation function depends on the delay closeness between the landmarks and the path end points.

5.1 Estimating bandwidth based on indirect paths' delays

Firstly, we estimate the end-to-end bandwidth of a direct path between two peers as equal to that of the shortest indirect path among the N indirect paths. More formally, the estimation function (Equation 1) is completed with the following expression of P_i:

$$P_i = \begin{cases} 1, & \text{if } RR_i = RR_{min} \\ 0, & \text{elsewhere} \end{cases} \tag{2}$$

where RR_i is the round trip delay of the indirect path that passes by the landmark L_i, and RR_{min} is that of the shortest indirect path among the N indirect paths.

We plot in Figure 3 the cumulative distribution function (CDF) of the bandwidth estimation accuracy which is calculated as the following:

$$accuracy = \frac{ABw_{estimated} - ABw_{measured}}{ABw_{measured}}. \tag{3}$$

The CDF is plotted over the 476 available direct paths. We observe that 25.72% of the estimations are accurate within 25% and 50.18% of the estimations are accurate within 50%. To check the correlation between the accuracy of the estimation and the difference in the delay between the direct and the shortest indirect path, we plot the histogram in Figure 4. For an estimation accuracy interval (on the x axis) of length 0.2, the y axis shows the ratio RR_{min}/R_d [3] averaged over all paths that have their corresponding accuracy within this interval. The figure does not show a clear trend, except the fact that for accuracies between -0.5 and 0.5, the ratio RR_{min}/R_d is small on average. This means that the pairs of peers that have their shortest indirect path delay close to the direct path delay behave better from bandwidth estimation standpoint.

Considering the shortest indirect path is not enough for providing accurate estimation since direct routing may use completely another path with close delay but different set of links. The accuracy could improve by considering more than one indirect path in the estimation function while assigning more weight to those having shorter delays. This consideration is mainly recommended when there are more than one indirect path having delays on the order of that of the shortest one.

Next, we consider all the N indirect paths in the bandwidth estimation function (Equation 1) with the following expression for the weight P_i:

$$P_i = \frac{C_i}{\sum_{i=1}^{N} C_i}, \quad \text{for } i = \{1, .., N\} \tag{4}$$

where,

$$C_i = \left(\frac{RR_{min}}{RR_i} \right)^{\alpha}, \tag{5}$$

and α is a positive real number.

We draw in Figure 5 the CDF of the estimation accuracy (Equation 3) for different values of α. The figure shows that when the α parameter increases, the estimation accuracy improves. This is expected since when $\alpha = 0$, the bandwidth component of all the indirect paths gets the same weight, and when α becomes large, the indirect paths having shorter delays, and hence better representation of the direct path, get more weight than those having larger delays. For $\alpha > 3$, we observe that the results become steady. This can be explained by the fact that the estimation becomes only dependent on the indirect paths having a delay on the order of that of the shortest indirect path. For $\alpha = 4$, the figure shows that 39.63% of the estimations are accurate within 25% and 68.36% of the estimations are accurate within 50%. These results prove clearly that the estimation is more accurate than the last case.

To show the correlation between the estimation accuracy and the difference in the delay between the direct and the indirect paths, we plot Figure 6 for the case $\alpha = 4$. For an estimation accuracy interval (on the x axis) of length 0.2, the y axis shows the sum $\sum_{i=1}^{N}(P_i \cdot RR_i)/R_d$, which is a weighted average of the ratio of the indirect paths' delays and the direct path delay (R_d). This sum is averaged over each interval of length 0.2. The figure shows a clear correlation between the two entities plotted on the x and y axis. This means that when

[3] R_d is the round trip delay of the direct path.

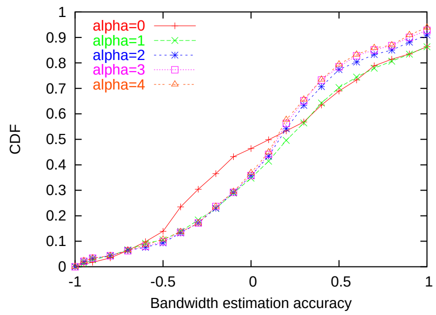

Fig. 5. Bandwidth estimation accuracy

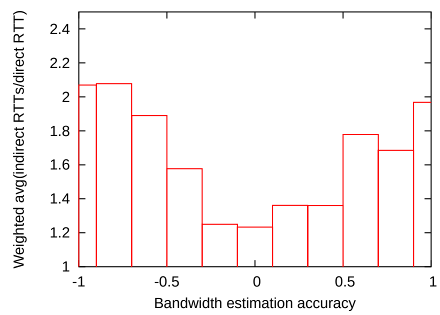

Fig. 6. Bandwidth estimation accuracy

some landmarks are located such that the delay of their correspondent indirect paths is close to that of the direct path, the estimation accuracy improves.

5.2 Estimating bandwidth based on the delay distance between landmarks and peers

To estimate the bandwidth of a direct path joining peers x and y, we now consider the indirect path whose landmark is the closest to one of the path extremities. We want to check if this indirect path is more representative of the direct path than the one having the smallest end-to-end delay. More formally, we consider for the bandwidth estimation function (Equation 1) the following form of the weight P_i:

$$P_i = \begin{cases} 1, & \text{if } R_i = R_{min} \\ \\ 0, & \text{elsewhere} \end{cases} \tag{6}$$

where,

$$R_i = min(R_{xi}, R_{yi}), \text{ [4]} \tag{7}$$

and

$$R_{min} = min_{i=1..N} R_i. \tag{8}$$

We plot in Figure 7 the CDF of the estimation accuracy. We observe that 28.57% of the estimations are accurate within 25% and 48.98% of the estimations are accurate within 50%. To evaluate the correlation between the estimation accuracy and the closeness of the landmark to one of the peers, we plot Figure 8. For an estimation accuracy interval (on the x axis) of length 0.2, the y axis gives R_{min} averaged over each x interval of length 0.2. The figure does not show a clear trend between these two entities. This could be an indication that it is not always sufficient to only consider the indirect path having the nearest landmark

[4] R_{xi} represents the round trip delay between the peer x and the landmark L_i.

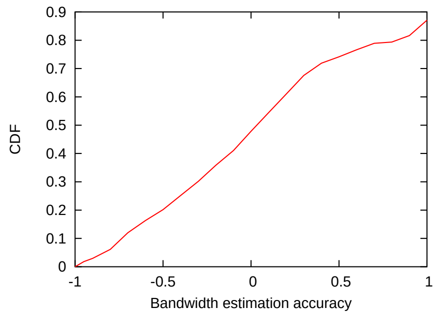

Fig. 7. Bandwidth estimation accuracy

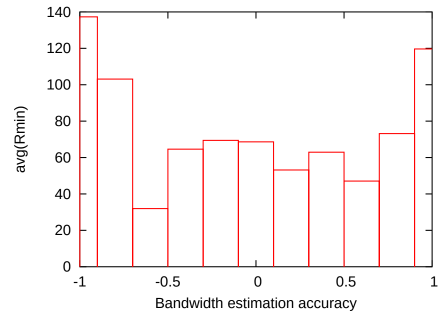

Fig. 8. Bandwidth estimation accuracy

to one of the extremities. We wonder if the consideration of more than one indirect path may improve the estimation accuracy after assigning more weight for those going through landmarks that are closer to the extremities.

Hence, for each pair of peers, we consider the N indirect paths in the bandwidth estimation model. We express the coefficients C_i as:

$$C_i = \left(\frac{R_{min}}{R_i}\right)^{\alpha}. \tag{9}$$

We recalculate the P_i function (Equation 4) and subsequently the estimation function (Equation 1) with these new coefficients C_i. Then, we plot in Figure 9 the CDF of the accuracy function (Equation 3) for all the bandwidth estimations and for different values of α. As before, when α increases, the indirect paths having landmarks close to one of the two peers get more weight. The results shown in the figure become stationary for $\alpha > 3$. This is because the bandwidth estimations become only dependent on the few indirect paths having landmarks close to one of the peers. The figure shows better results comparing to the previous cases studied; 56.54% of the estimations are accurate within 25% and 92.62% of the estimations are accurate within 50%.

To show the correlation between the estimation accuracy and the landmarks' closeness to the extremities, we plot Figure 10 for the case $\alpha = 4$. For an estimation accuracy interval (on the x axis) of length 0.2, the y axis shows $\sum_{i=1}^{N} P_i \cdot R_i$ averaged over the estimations inside the interval. The figure shows a clear correlation between the two entities in the x and y axis. This means that when some landmarks (among the N) are close to the path extremities, the estimation accuracy improves. Furthermore, it becomes better than the case where the estimation depends on the short indirect paths (see Figures 5 and 9). One interpretation is that sometimes the short indirect paths are not representative of the direct path. This is because the direct route provided by IP and those passing by the landmarks can be disjunct even if their delays are close to each other. On the other hand, the indirect paths going through landmarks that are

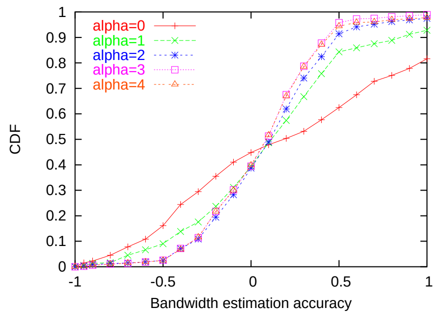

Fig. 9. Bandwidth estimation accuracy

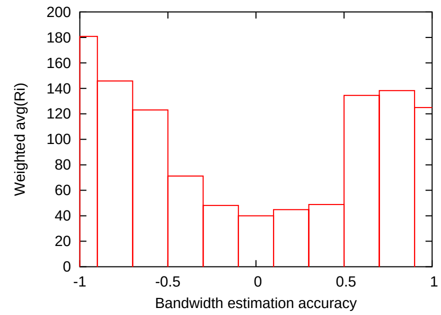

Fig. 10. Bandwidth estimation accuracy

close to path extremities, are more expected to provide better representation of the direct path.

6 Enhanced proximity perceived by the application

The proximity can be determined at the network-level by measuring a network parameter as the delay. We propose to determine it at the application-level by estimating some utility function that models the application quality such as the transfer time for the file sharing application. Within this new framework, peers are ranked from the standpoint of a certain peer in a decreasing order of the utility function. Close peers are those providing the best application quality independently of their network locality.

We denote by: (i) optimal proximity the one determined based on the utility function computed using the measured values of the network parameters, and by (i) approximate application-level proximity (ALP) the one determined based also on the utility function, but computed using the landmark-based estimated values of the network parameters. We aim to see whether the ALP proximity is a good approximation of the optimal one and how much the performance is different when compared to the basic delay proximity.

To this end, we consider a file transfer application over the TCP protocol. This case can be encountered in the emerging file sharing P2P applications or in the replicated web server context. Applications using TCP are known to form the majority of Internet traffic. For such applications, the optimal peer to select is the one allowing the transfer of the file within the shortest time. We call *latency* the transfer time. Since the impact of the bandwidth estimation is our main concern, we consider the case of large TCP transfers due to its sensitivity to this parameter [5].

In this section, we evaluate the degradation of the TCP latency when the delay proximity is used instead of the ALP proximity to perform the ranking of peers from the best to the worst. To predict the TCP transfer latency, we consider the function PTT (Predicted Transfer Time) that we compute in [5].

This function is the sum of a term that accounts for the slow start phase of TCP and another one that represents the congestion avoidance phase. We omit the window limitation caused by the receiver buffer and the loss rate along the network paths to allow a better understanding of the impact of the delay and the available bandwidth.

The degradation of TCP latency between the delay and the optimal proximity is computed as follows. Take a peer p and denote the peer having the rank r in the delay space by $p_d(r)$ (i.e., the peer having the r-th smallest RTT on its path to p). Denote by $p_o(r)$ the peer having a rank r with the optimal definition based on PTT and on the measured values of RTT and ABw. Let $PTT(x, y)$ denote the transfer latency between peer x and peer y. We define the $degradation_d$ at rank r as:

$$degradation_d(r) = \frac{PTT(p, p_d(r)) - PTT(p, p_o(r))}{PTT(p, p_o(r))}. \tag{10}$$

We repeat the same study for the ALP proximity. We denote the peer having the rank r in this space by $p_{ap}(r)$. This peer has, on its path with p, the r-th smallest PTT which is computed based on estimated values of ABw. We use the measured values of RTT instead of the landmark-based estimated ones in order to focus on the impact of our bandwidth estimation approach on the application performance. Thus, the $degradation_{ap}$ at rank r is:

$$degradation_{ap}(r) = \frac{PTT(p, p_{ap}(r)) - PTT(p, p_o(r))}{PTT(p, p_o(r))}. \tag{11}$$

Using the last model of Section 5.2 (i.e., the case where all indirect paths are considered and the closeness between the landmarks and the peers is the criterion) and for $\alpha = 4$, we infer the available bandwidth for the paths between a peer p (each of the 14 peers) and the 34 other peers to determine the latency degradation for a large file transfer ($S = 10MB$) on each path. Then, we average all degradation values at rank r over the 14 peers. This study allows to evaluate how well ranking peers based on the delay proximity and on the ALP proximity performs on average at the application level with respect to the optimal case.

We plot in Figure 11 the transfer time degradation function of the rank r for the delay and the ALP proximity. The closest 10 peers are considered. The figure shows that the degradation is much larger when the proximity is based on the delay and it does not exceed the 40% when using the ALP proximity. Long transfers are more sensitive to the bandwidth (i.e., bandwidth greedy) and since the bandwidth is uncorrelated with the delay (as obtained in Section 3), considering the delay alone for proximity characterization is far from being optimal. Situation improves considerably when bandwidth estimations are considered.

7 Conclusion

We introduce in this paper a new notion of proximity that accounts for path characteristics and application requirements. We motivate the need for this notion by showing that the proximity in the delay space does not automatically

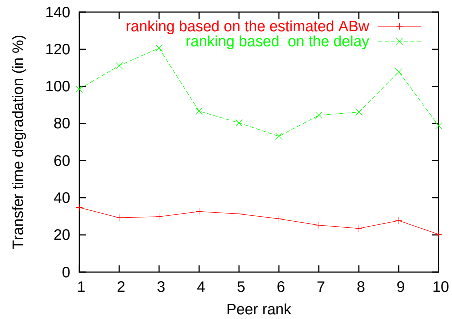

Fig. 11. Transfer time degradation

lead to a proximity in the bandwidth space. The proximity needs to be defined as a function of the metrics impacting the application performance. Then, we extend the landmark approach used for estimating the delay to infer scalably the bandwidth among peers. Finally, we show that the proximity determined using our bandwidth estimation model provides much lower quality degradation than that obtained using the delay proximity for file transfer applications.

References

1. B. Wong, A. Silvkins, and E. G. Sirer: Meridian: A Lightweight Network Location Service without Virtual Coordinates, Sigcomm'05, 2005.
2. F. Dabek, R. Cox, F. Kaashoek, and R. Morris: Vivaldi: A Decentralized Network Coordinate System, Sigcomm'04, 2004.
3. B. Gueye, A. Ziviani, M. Crovella, and S. Fdida: Constraint-Based Geolocation of Internet Hosts, IMC'04, 2004.
4. V. Padmanabhan and L. Subramanian: An Investigation of Geographic Mapping Techniques for Internet Hosts, SIGCOMM'01, 2001.
5. M. Malli, C. Barakat, and W. Dabbous: An Efficient Approach for Content Delivery in Overlay Networks, CCNC'05, 2005.
6. J. Navratil, L. Cottrell: ABwE: A Pratical Approach to Available Bandwidth Estimation, PAM'03, April 2003.
7. J. Navratil, L. Cottrell: available at *http://www-iepm.slac.stanford.edu/tools/abing/*, 2004.
8. J. Navratil: available at *http://www.slac.stanford.edu/ jiri/PLANET*, 2004.
9. An open, distributed platform for developing, deploying and accessing planetary-scale network services, see *http://www.planet-lab.org/*.
10. E. Ng and H. Zhang: Predicting Internet network distance with coordinates-based approaches, IEEE Infocom, 2002.
11. L. Tang, and M. Crovella: Virtual Landmarks for the Internet, IMC'03, 2003.
12. Y. Shavitt, and T. Tankel: Big-bang simulation for embedding network distances in euclidean space, IEEE Infocom, 2004.
13. L. Peterson, V. Pai, N. Spring, and A. Bavier: Using PlanetLab for Network Research: Myths, Realities, and Best Practices, technical report, 2005.
14. S. Ratnasamy and M. Handly and R. Karp and S. Shenker: Topologically-Aware Overlay Construction and Server Selection, IEEE Infocom, 2002.