An Enhanced Scalable Proximity Model

Mohammad Malli, Chadi Barakat, Walid Dabbous
INRIA-Sophia Antipolis, France

In overlay networks, the proximity of peers is almost determined at the network-level by measuring 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. The challenge is to infer the network parameters, included in the utility functions, in an easy scalable way.

It consists in inferring the network parameters among a large number of peers without achieving a mesh-based measurement. Many approaches have been proposed for estimating the end-to-end delay among peers from a set of partially observed measurements. Most of them are based on embedding network distances solutions [1].

In this work, we investigate the bandwidth estimation problem and it is impact on the quality of service of the file sharing application. This is completely different from the problem of the delay estimation because path available bandwidth is determined by the bottleneck link that may appear anywhere along the path.

Our scheme consists in inferring the bandwidth between any pair of peers based on their bandwidth vectors. A peer obtains its bandwidth vector by measuring the direct and reverse bandwidth on its path to a set of intermediate nodes. In the rest of paper, we refer to these nodes by the term *landmarks*. Hence, the scheme is scalable since the system overhead is linear with the number of peers in the system. Also, it is easy to implement since peers do 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. We notice that N indirect paths (i.e., N being the number of landmarks) are assigned to each direct path.

We estimate the end-to-end bandwidth of a path joining two peers using the following class of linear functions:

$$EB = \sum_{i=1}^{N} P_i \cdot BB_i, \qquad (1)$$

where BB_i is the bandwidth of the indirect path that passes by the landmark L_i, and P_i is the normalized weight (i.e., $\sum_{i=1}^{N} P_i = 1$) assigned to this indirect path according to the location of its corresponding landmark with respect to the two peers. The idea behind this definition of the estimation function is as follows. We consider that the direct path shares the same tightest link with the indirect path that passes by the landmark L_i with some probability P_i. This probability depends on the location of the corresponding landmark with respect to the direct path or to one of its end points. By varying the expression of the probability 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.

We have mainly studied these two cases: (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. In our experiments, the second class of functions provides the best estimation accuracy. Next, we present only this case to fit on the paper size.

Hence, for each pair of peers, we consider the N indirect paths in the bandwidth estimation function (Equation 1) after assigning more weight for those going through landmarks that are closer to the two peers under consideration. Thus, we express the weight P_i as:

$$P_i = \frac{C_i}{\sum_{i=1}^{N} C_i}, \quad \text{for } i = \{1, .., N\}, \qquad (2)$$

where, $C_i = (R_{min}/R_i)^{\alpha}$, $R_i = min(R_{xi}, R_{yi})$, R_{xi} represents the round trip delay between the peer x and the landmark L_i, $R_{min} = min_{i=1..N} R_i$, and α is a positive real number. When α parameter increases, the indirect paths having landmarks close to one of the two peers get more weight.

To evaluate the bandwidth estimation accuracy, we consider the following real scenario conducted over Planetlab. We take 8 Planetlab nodes distributed in different European countries as landmarks. We also take 14 Planetlab nodes completely distributed in Europe as peers. We wonder whether placing the landmarks in Europe leads to estimate accurately the bandwidth on the path between an European peer and any other peer. Therefore, each of the 14 European 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.

We draw in Figure 1 the CDF of the accuracy function, which is calculated as $(ABw_{estimated} - ABw_{measured})/ABw_{measured}$, for different values of

1-4244-0476-2/06/$20.00 ©2006 IEEE.

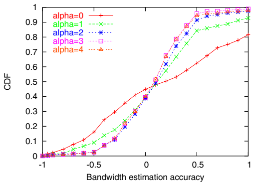

Fig. 1. Estimation accuracy

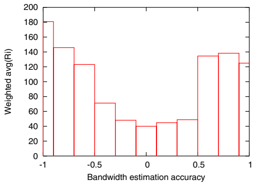

Fig. 2. Estimation accuracy

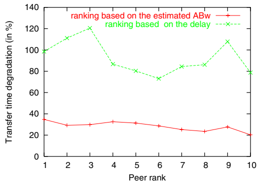

Fig. 3. Transfer time degradation

α. The figure shows that when α parameter increases, the estimation accuracy improves. This is expected since when $\alpha = 0$, the bandwidth contribution of all indirect paths gets the same weight, and when α becomes large, the indirect paths having landmarks close to one of the two peers, and hence better representation of the direct path, get more weight than those having landmarks farer from the end points. 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 closer landmarks to the end points. For $\alpha = 4$, the figure shows that 56.54% of the estimations are accurate within 25% and 92.62% of the estimations are accurate within 50%.

Figure 2 shows the correlation between the estimation accuracy and the landmarks' closeness to the two peers 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 laying inside the interval. We can observe a clear correlation between the two entities on the x and y axis. This means that when some landmarks (among the N) are close to the path extremities, the estimation accuracy improves.

In the rest of the paper, we study whether the application-level proximity (ALP)[1] is a good approximation of the optimal proximity[2] 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. For such applications, the optimal peer to select is the one allowing the transfer of the file within the shortest 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 [2].

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 [2]. 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 $degradation_d(r) = (PTT(p, p_d(r)) - PTT(p, p_o(r)))/PTT(p, p_o(r))$.

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. Then, the degradation of TCP latency between the ALP proximity and the optimal proximity is computed at rank r as $degradation_{ap}(r) = (PTT(p, p_{ap}(r)) - PTT(p, p_o(r)))/PTT(p, p_o(r))$.

Using the estimation model previously presented 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 3 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. Thus, considering the delay alone for proximity characterization is far from being optimal for large file transfer applications. Situation improves considerably when bandwidth estimations are considered.

Our future work aims at determining the impact of the number of landmarks and their distribution on the bandwidth estimation accuracy. On another venue, we are looking for how one can profit from bandwidth measurements to the landmarks to optimize overlay construction.

REFERENCES

[1] E. Lua, T. Griffin, M. Pias, H. Zheng, and J. Crowcroft, On the Accuracy of Embeddings for Internet Coordinate Systems, IMC'05, 2005.

[2] M. Malli, C. Barakat, and W. Dabbous: An Efficient Approach for Content Delivery in Overlay Networks, IEEE CCNC, 2005.

---

[1] ALP is the application-level proximity determined based on the utility function computed using the inferred values of the network parameters.

[2] The optimal proximity is the one determined based on the utility function computed using the measured values of the network parameters.