A Hierarchical Approach to Internet Distance Prediction

Rongmei Zhang[1], Y. Charlie Hu[1], Xiaojun Lin[1], and Sonia Fahmy[2]
[1] School of Electrical and Computer Engineering
[2] Department of Computer Science
Purdue University, West Lafayette, IN 47907
{rongmei, ychu, linx, fahmy}@purdue.edu

Abstract

Internet distance prediction gives pair-wise latency information with limited measurements. Recent studies have revealed that the quality of existing prediction mechanisms from the application perspective is short of satisfactory. In this paper, we explore the root causes and remedies for this problem. Our experience with different landmark selection schemes shows that although selecting nearby landmarks can increase the prediction accuracy for short distances, it can cause the prediction accuracy for longer distances to degrade. Such uneven prediction quality significantly impacts application performance. Instead of trying to select the landmark nodes in some "intelligent" fashion, we propose a hierarchical prediction approach with straightforward landmark selection. Hierarchical prediction utilizes multiple coordinate sets at multiple distance scales, with the "right" scale being chosen for prediction each time. Experiments with Internet measurement datasets show that this hierarchical approach is extremely promising for increasing the accuracy of network distance prediction.

1. Introduction

Internet distance (latency) prediction estimates the pairwise ($O(N^2)$) distances between N nodes, using only $O(N)$ measurements. In recent years, several prediction mechanisms have been proposed, e.g., [12, 4, 5, 11]. Most of these mechanisms assign a coordinate to a node in order to characterize its approximate location in the Internet; the distance between two nodes can be estimated by applying a distance function over their coordinates. For the purpose of determining its coordinate, a node usually selects a number of other nodes as reference points (often referred to as "landmarks"). The coordinate is then computed by minimizing the difference between the estimated distances and the measured distances to these landmarks.

The accuracy (or inaccuracy) of a prediction scheme has often been measured by the relative prediction error *over all distance ranges* [12]. When this metric is used to evaluate the quality of the prediction, existing prediction algorithms have appeared to achieve reasonably high prediction accuracy. However, when the prediction errors are broken down to each distance range, there typically exists a significant disparity in the prediction accuracy for different distance ranges. Evaluation of various prediction algorithms proposed in the literature suggests that these algorithms tend to over-estimate short distances and under-estimate long distances.

Several approaches have been proposed in the literature to improve the distance prediction accuracy, including increasing the dimension of the coordinate space or the number of landmark nodes. However, it has been observed that, once beyond a certain threshold, the prediction cannot benefit from higher dimensions or a larger number of landmarks [12, 4]. Another approach that has been proposed is to "intelligently" determine the placement of the landmark nodes. Previous studies do indicate that the prediction accuracy can be very sensitive to the placement of the landmark nodes [12]. In [4], the authors have explored various landmark selection schemes, and recommended a *hybrid landmark selection scheme* that uses a combination of randomly selected landmark nodes and nearby landmark nodes. However, using both synthetic and Internet measurement datasets, we observe that although the hybrid landmark selection scheme can improve the prediction accuracy for short distances, other distance ranges may suffer degraded prediction accuracy. The potential interference between short and long distances can cause the application to experience degraded performance. In Section 3, we propose a new metric, *reverse range accuracy*, which captures the impact of this interference between distance ranges when prediction is used. Our experiments [23] show that using a hybrid selection of landmarks can notably degrade the reverse range accuracy.

The objective of this work is to investigate how to improve the prediction accuracy of short distances without degrading the accuracy of long distances. Observing that none

of the distance prediction schemes that we have evaluated can simultaneously improve the prediction accuracy of both short and long distances, we conjecture that there may be some fundamental reasons why accurately predicting both short and long distances is difficult [23]. In particular, we study the impact of the range and distribution of the distances on the prediction accuracy and find that a narrower distance range contributes to higher prediction accuracy.

Motivated by the above observation, we resort to alternative approaches to improving Internet distance prediction. In Section 5, we propose a hierarchical distance prediction approach that computes *multiple* sets of coordinates for each node using multiple landmark sets, where each set of coordinates and landmarks correspond to a different distance scale. For instance, a node can use one coordinate for estimating the distance to a distant node, and a second coordinate for estimating the distance to a close node. Our main idea is that if we can find one set of coordinates that is more accurate for short distances, and another set of coordinates that is more accurate for long distances, then depending on the distance range of interest, we can choose the right set of coordinates to achieve higher overall prediction accuracy. Our evaluation demonstrates that the hierarchical approach is very effective in improving the prediction accuracy of short distances at little or no cost in the prediction accuracy of long distances.

2. Related Work

Several approaches have been proposed for network distance prediction. The distance between two nodes can be inferred from the measured distances to the landmarks, e.g., based on the assumption of the triangle inequality [8], or the order of the distances to the landmarks [15]. In [9, 18], the distances to the landmarks are assigned as the coordinate of the node in a multi-dimensional space. In GNP [12, 13], node coordinates are computed using the Simplex Downhill method. Lighthouse [14] derives node coordinates by solving systems of linear equations. Mithos [20] and Vivaldi [5] use the spring relaxation technique to calculate node coordinates. In [16, 17], the Big-Bang simulation method is used for embedding network distances in a multidimensional space. IDES (Internet Distance Estimation Service) [11] exploits matrix factorization, i.e., Singular Value Decomposition (SVD) or Non-negative Matrix Factorization (NMF), to compute an incoming and an outgoing coordinate for each node. In [4], different strategies for choosing landmarks are studied, including using random nodes, closest nodes, and a hybrid of both. Choosing nearby nodes as landmarks is shown to improve the prediction accuracy for short distances. In addition to these coordinate-based prediction mechanisms, there have been proposals for distance prediction based on a measurement infrastructure [6, 19, 3, 21, 22].

3. Evaluation Methodology

In this section, we describe the evaluation methodology that we use throughout the paper.

Internet Distance Datasets Our evaluation uses 3 latency datasets from real-world measurements. The P2PSim dataset is from the MIT P2PSim project [2]. This dataset contains the pair-wise RTTs between 1740 nodes measured using the KING method [7]. The Meridian dataset is from the Cornell Meridian project [21]. It measures the pair-wise RTTs between 2500 nodes using the KING method [7]. The AMP dataset is from the NLANR Active Measurement Project [1]. The 110 nodes in this dataset are connected to high speed networks.

The average RTTs in these three datasets are about 182 ms, 75 ms, and 55 ms respectively. The maximum RTTs in the P2PSim and the Meridian datasets are around 800 ms and 996 ms respectively, while the AMP dataset has a maximum RTT of 373 ms. We note that all three datasets exhibit high concentration over small distance ranges. For the P2PSim dataset, more than 85% of links are below 300 ms; for the Meridian and AMP datasets, links of up to 100 ms count for over 80% and 90% of all the links respectively. Links of medium lengths are the majority, while short links are not as abundant. For instance, in the P2PSim dataset, only about 1.6% of links are within 20 ms (roughly the typical RTT within a local region), and about 6.9% are within 40 ms (approximate RTT of halfway across the continental US).

Distance Prediction Algorithms In most of our experiments, we study a GNP-style prediction algorithm, and evaluate options for each node to select its own landmark nodes. The coordinate is computed using the simplex-downhill algorithm as in [12, 4]. As in [4], we assume an 8-dimensional coordinate space with 16 landmarks used for prediction. Our previous study [24] shows that there are no major performance differences between distance prediction schemes that have been proposed so far from the perspective of the applications. We have repeated a subset of our experiments with IDES-style prediction and the results were consistent with the results discussed here.

Evaluation Metrics The prediction accuracy of a network distance prediction mechanism is often measured by the **relative prediction error**:

$$\frac{abs(predicted\ distance - measured\ distance)}{measured\ distance}.$$

In this paper, we propose a new metric, **Range Accuracy**, to gauge the ability to correctly identify links that are inside a particular distance range r (specified by a lower

and upper threshold) using the prediction algorithm. Consider a distance range r, and a distance prediction algorithm that predicts that all links falling within r comprise a set $PredictedLinks$. Assume that the set $MeasuredLinks$ comprises all links that fall within r if we measure actual distances. We define the **Range Accuracy** as the fraction of correctly identified links among all those actually within the given range r, i.e., $\frac{|MeasuredLinks \cap PredictedLinks|}{|MeasuredLinks|}$, where $|X|$ denotes set X cardinality. Conversely, we define the **Reverse Range Accuracy** as the fraction of predicted links that are indeed within the given range r, i.e., $\frac{|MeasuredLinks \cap PredictedLinks|}{|PredictedLinks|}$. While the **Range Accuracy** indicates the prediction accuracy for those links within the given range, the **Reverse Range Accuracy** reflects the interference between distance ranges caused by prediction inaccuracy. The two metrics give the converse of false negatives (links mis-predicted to fall outside r) and converse of false positives (links mis-predicted to fall within r) respectively.

In [10], two additional metrics, i.e., Relative Rank Loss (rrl) and Closest Neighbor Loss (cnl), were proposed to capture the prediction quality as perceived by the user. Due to space limits, we focus on the range accuracy and reverse range accuracy metrics in this paper. Evaluation results with the rrl and cnl metrics can be found in an extended version [23].

4. Impact of Distance Range and Distribution on Prediction Accuracy

In this section, we study the impact of the range and distribution of the distances on prediction accuracy. Our hypothesis is that *the narrower the distance range, the higher the prediction accuracy*. The intuition is that it is easier to embed the nodes into a multi-dimensional space when the pair-wise distances between them are concentrated within a small range. Conversely, it is difficult to position a node accurately with regard to both relatively nearby nodes and distant outliers. Moreover, a node coordinate is oftentimes computed by minimizing the overall prediction error (of all node pairs or from a node to the landmarks), e.g., by using the sum of squares error function. Therefore, a higher concentration within some distance range tends to result in higher prediction accuracy over the range. Since randomly selected landmarks provide a random sample of the distances to be predicted, the prediction accuracy is likely to follow the distance density distribution, with high-density ranges having higher prediction accuracy.

4.1. Experiments with Internet Distance Datasets

First, we evaluate the hypothesis using the three distance datasets. We assume a set of landmark nodes randomly chosen from each dataset, which serve as reference points for all nodes during prediction. Fig. 1(a) shows the CDF of the relative prediction error. The AMP data has higher overall prediction accuracy than the P2PSim and Meridian datasets. Fig. 1(b) depicts the relative prediction error over the distance spectrum: short links apparently suffer from higher relative prediction error. While all three datasets experience similar prediction error for medium and long distances, e.g., distances beyond 50 ms, the AMP dataset has notably lower error for short links.

Fig. 1(c) shows the range accuracy. In this experiment, we evaluate three distance ranges (0,20 ms], (0,40 ms] and (0,80 ms]. As shorter distances are more difficult to predict accurately, the range accuracy degrades for lower distance thresholds. In fact, the accuracy for selecting the links within 20 ms is below 50% for both the P2PSim and Meridian datasets. Note that the reverse range inaccuracy is also significant (Fig. 1(d)). For instance, among all the links predicted to be within 40 ms in the Meridian dataset, only 36% are in fact of more than 40 ms. This can be attributed to the interference from mis-predicted (under-estimated) longer links. Overall, the AMP dataset has significantly better accuracy than the other two datasets.

Recall that while all three datasets exhibit roughly similar distance density distributions, the AMP dataset has a much narrower distance range, with a maximum distance of 373 ms. The above results confirm our hypothesis that a narrow distance range contributes to high prediction accuracy. We will omit the experiments from the AMP dataset beyond this subsection, since the prediction quality for this dataset is good with most prediction schemes. Due to limited space, we will also omit results from the Meridian dataset, which can be found in the extended version [23].

4.2. Experiments with Synthetic Datasets

In this subsection, we conduct experiments with synthetic datasets to further evaluate our hypothesis. In order to control both the range and distribution of the distances to be predicted, we generate these synthetic datasets as follows. First, a set of N points is generated with random locations in a multi-dimensional space. Then, the pair-wise (Euclidean) distances between the N points are computed to generate an $N \times N$ distance matrix, which completely conforms to the triangular inequality. Since Internet distances have been observed to exhibit triangular inequality violations (TIV), we introduce errors to the distance matrix, and the error level is controlled by a parameter δ. The range of the synthetic distances can be bounded by given lower and upper thresholds. In our first dataset, $N = 1000$, and $\delta = 30\%$, i.e., on average, the error magnitude is 30% of the original Euclidean distance. For comparison, we generate a second dataset with a much smaller upper distance threshold of only 50 ms. In order to closely match the distance
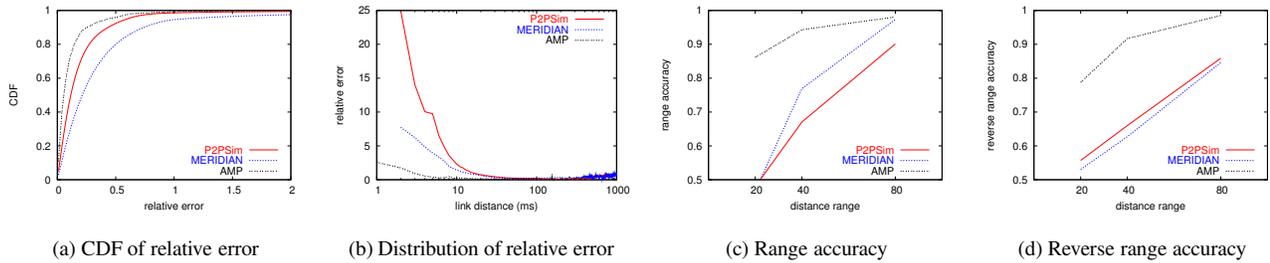
| (a) CDF of relative error | (b) Distribution of relative error | (c) Range accuracy | (d) Reverse range accuracy |

Figure 1. Prediction accuracy with random global landmark selection

distribution of the second dataset to the first dataset (within 50 ms), we set $N = 600, \delta = 30\%$ in the second dataset. The density distributions of the two synthetic datasets are shown in Fig. 2(a).

Fig. 2(b) shows the CDF of the relative prediction error. The mean relative prediction errors are 0.40 and 0.27 for the two generated datasets respectively. If we consider solely the links within 50 ms in the first dataset, the mean relative prediction error is as high as 0.45. Clearly, a narrow distance range helps to improve the prediction accuracy. These results are consistent with the observation made in [10] that "less information gives better accuracy results." In fact, our studies reveal that the reason behind this phenomenon is closely tied to the distance range in question. We can also observe the association between prediction accuracy and the distribution of distances from Fig. 2(c), which shows the absolute prediction error, i.e., $abs(predicted\ distance - measured\ distance)$.

4.3. Impact of Landmark Selection

We have also evaluated the impact of adopting "intelligent" landmark selection schemes discussed in [4], including random landmark selection, closest landmark selection, and a hybrid of random and closest landmarks. In contrast to the experiments presented above, where all nodes use the same set of "global" landmarks, each node selects its own landmarks from the nodes with known coordinates. The results (see [23]) from using the three Internet latency measurement datasets can be summarized as follows. The random selection scheme can achieve comparable prediction quality to the random global landmark selection. Using closest nodes as landmarks can improve the prediction accuracy for the shortest links, while the majority of links (of medium or long length) experience degraded accuracy.

In addition to the above experiments, our analytical results (see [23]) also suggest that it is fundamentally difficult to accurately predict both short and long distance using a single coordinate per node.

5. Improving Prediction Accuracy: A Hierarchical Approach

In this section, we propose a new approach which exploits multiple coordinates for the same node, each positioning the node at a different distance scale. This can be considered as a hierarchy of multiple levels: level-0 is the "global" scale, and each subsequent level-i covers a shrinking distance range.

5.1. Leveraging A Shared Hierarchy

We first explore constructing a landmark hierarchy that is shared by all nodes for prediction. Specifically, a number of landmark nodes form a hierarchy through recursive clustering. Each cluster consists of landmark nodes that are close to each other. The cluster heads from lower-level clusters are members of the next high-level cluster. The hierarchy levels represent increasingly finer distance scales, starting from the top-level which represents the "global" scale.

A node maintains separate coordinates for different levels in the hierarchy. At the top level (level-0), each node computes a coordinate with reference to level-0 landmarks. At level-1, there are L landmark clusters. A node can choose one or more clusters based on their distance proximity to compute one or more level-1 coordinates. The number of coordinates at level-i is designated by a parameter α_i. A node can infer the proximity of a landmark cluster by measuring the distance to the cluster head. Alternatively, a node can also determine the closeness of a landmark cluster based on the maximum or average distance to the cluster members.

Prediction proceeds in a bottom-up fashion. Starting from the bottom level, it is first determined whether two nodes, say A and B, are both close to some cluster at the current level, i.e., if both maintain a coordinate relative to the same landmark cluster. If so, this implies that nodes A and B are relatively close to each other, and therefore the distance can be estimated using the corresponding coordinates at current level. Otherwise, the next higher level is tested, until the top level, i.e., the global level is reached. It is possible that nodes A and B share more than one landmark set at level-i. In this case, the average of the predicted

(a) Density distribution of RTT   (b) CDF of relative error   (c) Distribution of absolute error
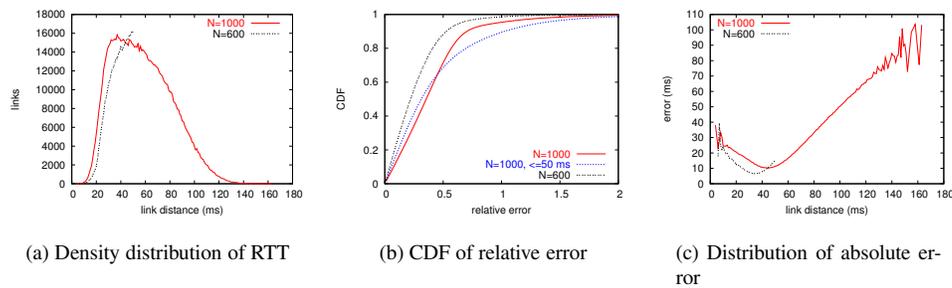
Figure 2. Impact of distance range and distribution

values with reference to all the shared landmark sets can be taken.

5.1.1 Evaluation

To experimentally evaluate the above approach, we use a 2-level hierarchy. The bottom level (level-1) comprises 16 clusters, each containing 16 landmarks. The representatives from the 16 level-1 clusters form the top level (level-0). A node maintains 4 coordinates with reference to the closest 4 landmark sets at the bottom level, i.e., $\alpha_1 = 4$.

Landmark Hierarchy Construction   In our experiments, we use a simple heuristic to build a landmark hierarchy for each of the P2PSim and Meridian datasets. First, β nodes are selected from the dataset, where $\beta > 16$. The β nodes are the seeds to grow the clusters. They are randomly chosen and are well-distributed, i.e., the pair-wise distances between the β nodes are above a given threshold γ. A cluster is formed by adding one node per round. In each round, the seeds take random turns to choose the closest nodes to themselves from the remaining nodes in the datasets. 16 out of the β clusters are then selected to form the landmark hierarchy, based on the "quality" of the candidate clusters. We define the quality by the average pair-wise distance between the cluster members, and randomly select 16 clusters of average pair-wise distances below a given threshold θ. In our experiments, $\gamma = 150$ ms, $\theta = 100$ ms for the P2PSim dataset, and $\gamma = 100$ ms, $\theta = 60$ ms for the Meridian dataset.

Evaluation Results   We now present the results of evaluating the shared hierarchical prediction scheme. The label "random global" represents the results using the top-level landmarks, while "hierarchy shared" represents the results using the 2-level landmark hierarchy. We also compare with the hybrid landmark selection scheme, which selects 4 nearby nodes and 12 random nodes as landmarks.

The evaluation results for the P2PSim dataset are reported in Fig. 3(a)-Fig. 3(d). Fig. 3(a) shows that using the 2-level landmark hierarchy improves the overall prediction accuracy. Fig. 3(b) shows the average relative prediction error for links of various distances. Compared to

using the top-level coordinates alone, the hierarchical approach improves the prediction quality for links of up to about 300 ms, while longer links can experience slightly degraded prediction accuracy. This is because short distances are more likely to be estimated using the bottom-level coordinates, and hence to experience higher prediction accuracy. The hybrid landmark selection scheme also achieves high prediction accuracy for short links. However, the overall prediction quality cannot match the hierarchical prediction scheme, due to the notably worse prediction accuracy for medium and long links, i.e., links of delay above 150 ms.

Fig. 3(c) and Fig. 3(d) indicate that both the range accuracy and the reverse range accuracy are significantly improved by using the hierarchical prediction. This reflects both lower false negative rate and lower false positive rate in selecting short distances. On the other hand, although the hybrid landmark selection scheme achieves higher range accuracy, the reverse range accuracy decreases. This can be explained by the interference from longer links due to the higher prediction error beyond short distances.

Note that the performance of the hierarchical prediction varies with the landmark placement at the top level (i.e., level-0). Choosing a different set of landmarks at the top level might produce different results. However, adopting the landmark hierarchy always out-performs using one set of landmarks (i.e., at the top level).

5.2. Hierarchical Prediction with Flexible Landmark Selection

Instead of building a shared landmark hierarchy, the second hierarchical prediction scheme allows each node to choose its own landmarks. The "hierarchy" is specified by the radius R_i of each level-i. For instance, level-1 is defined by $R_1 = 100$ ms. Given a node B, node A first determines an approximate distance range where node B is likely to be located, and then uses their corresponding coordinates at the estimated level for predicting the distance. If node A predicts node B to be within 100 ms, their level-1 coordinates are then used to compute the distance. At each level-i, a node only needs to choose a set of other nodes as landmarks that are within (or close to) R_i to itself. Therefore, level-i

5

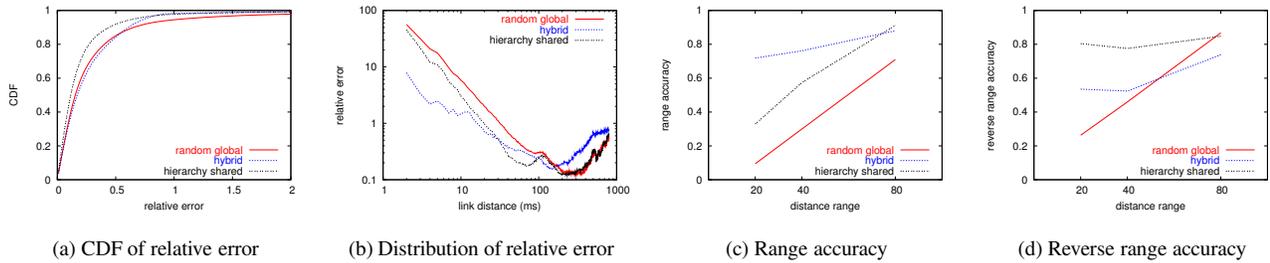| (a) CDF of relative error | (b) Distribution of relative error | (c) Range accuracy | (d) Reverse range accuracy |

Figure 3. Prediction accuracy with a shared landmark hierarchy (P2PSim). The random global scheme uses the same set of landmarks as the level-0 of the 2-level hierarchy, which are the cluster heads from the level-1.

coordinates are more accurate for predicting distances that are within R_i, and less accurate for predicting distances that are beyond R_i.

The main challenge for this hierarchical approach lies in accurately estimating the distance range between two nodes, and then selecting the right level to be used for predicting the distance. In this paper, we leverage the triangular inequality existing in the Internet for the purpose of distance range estimation. Recent studies have found that triangular inequality is violated with a typical possibility of 10-20% [25]. Therefore, we can expect it to hold for the majority of the cases. The scheme assumes a set of nodes is distributed over the network. For clarity, we call these nodes "anchors" (specialized landmarks). The anchors should be able to achieve good coverage of the network, i.e., most nodes should be within relatively small distances to at least one of these anchors. Assuming the triangular inequality stands, if node A and B are both within $R_i/2$ to some anchor C_j, the distance between node A and B is upper bounded by R_i.

5.2.1 Evaluation

We assume a hierarchy of three levels: level-0 as the global level, level-1 covering distances of up to 40 ms, and level-2 covering distances of up to 20 ms.

Anchor Selection In our experiment, we use an offline heuristic to identify a set of nodes that provide good coverage for each dataset. The offline heuristic selects a node from the dataset each time, such that the node can add to the current covered node set at least σ more nodes. In our experiment, $\sigma = 2$: each selected anchor covers at least 2 more nodes. The anchor selection and the achievable coverage are designated by the radius. In our experiment, the radius is set to to half of the range of the lowest level (level-2), i.e., $R_2/2$. The selected anchors can be shown to provide good coverage for level-1 (see Table 1).

We evaluate the accuracy of using the selected anchors for range estimation. Specifically, we measure the false positive rate and false negative rate at level-2 (with radius 10 ms) and level-1 (with radius 20 ms), and the results are

reported in Table 1. Both datasets experience high false negative rates, while the false positive rates stay relatively low. For instance, over 66% of the links below 20 ms in the P2PSim dataset are estimated to be over 20 ms based on measurements to the anchors; these links will not be able to benefit from the higher prediction accuracy of the level-2 prediction.

There are two factors that contribute to the high false negative rates. First, the selected anchors are unable to provide perfect coverage for the dataset. The higher false negative rate for the P2PSim dataset can be explained by the lower coverage rate, due to the low density of short distances in the dataset. Second, the range estimation accuracy based on measurements to the anchor nodes also depends upon conformance to the triangular inequality. The lower false positive rate for the P2PSim dataset reflects higher degree of conformance to the triangular inequality.

Anchors can also serve as landmarks for prediction. For instance, if node A is within distance $R_2/2$ to two anchors C_1 and C_2, the anchors can be selected as landmarks for prediction at level-2. In our experiment, the anchors also provide bootstrapping capability by computing their coordinates first. In addition, 16 anchors are randomly selected to serve as landmarks at the global level (i.e., top level of of the hierarchy).

Evaluation Results Our experiments evaluate the hierarchical prediction approach in comparison to two baseline schemes: random global landmark selection and hybrid landmark selection. The random global landmark selection scheme uses the same set of landmarks as the global level of the hierarchical approach. The hybrid random selection is configured with 4 (out of 16) nearby landmarks.

Fig. 4(a) depicts the CDF of relative prediction error for the entire P2PSim dataset: the hierarchical approach experiences similar prediction quality as the random global landmark selection scheme. Although using the hybrid landmark selection scheme improves the prediction accuracy for short links (Fig. 4(b)), longer links suffer decreased prediction accuracy, causing the overall prediction quality to degrade. Fig. 4(c) shows that both the hierarchical scheme and

Table 1. Anchor Selection and Range Estimation Accuracy

	anchors	level-2			level-1		
		coverage	false positive	false negative	coverage	false positive	false negative
P2PSim	111	961	0.089%	66%	1359	0.45%	55%
Meridian	77	2022	6.8%	45%	2307	17%	30%

the hybrid landmark selection scheme can improve the ability to correctly identify the links within 20 ms, 40 ms, and 80 ms respectively. However, the hybrid landmark selection scheme causes the reverse range accuracy to degrade, due to the higher prediction error for longer links. Lower reverse range accuracy corresponds to higher false positive rate in range estimation.

The performance of the flexible hierarchical prediction approach also varies with the landmark selection at the global level (level-0). However, we expect that the hierarchical approach can always improve the prediction quality over short distances without severe degradation over other distance ranges.

Impact of Range Estimation Inaccuracy In order to illustrate the impact of the range estimation accuracy on the performance of the hierarchical prediction mechanism, we utilize an "oracle" that is able to precisely determine the right range (e.g., level) for the target distance. Fig. 5(a)-Fig. 5(c) report the results of applying the oracle with the anchor-based hierarchical scheme on the P2PSim dataset. We observe improved prediction quality, which in turn benefits the application performance. This experiment indicates that an accurate range estimation scheme is key to the success of the hierarchical prediction mechanism and anchor-based range estimation is a promising technique.

5.3. Practical Considerations

In this section, we have presented two schemes for hierarchical distance prediction. The first leverages a shared landmark hierarchy, and thus landmark nodes may sustain a high measurement load. In the second scheme, the measurement load is distributed over all the nodes as each node chooses its own landmarks. In our experiments, we have adopted two heuristics for building the landmark hierarchy and for selecting the anchors from the nodes in the datasets respectively. In practice, both the landmarks and anchors can be determined according to the inherent hierarchy of the Internet, with each sub-network being deployed with its own landmarks or anchors.

Since each node maintains multiple coordinates, the hierarchical prediction approach incurs a higher overhead than non-hierarchical prediction. In the shared hierarchical scheme, the overhead is designated by the number of coordinates (i.e., α_i) that a node maintains at each level-i. In the second hierarchical scheme with flexible landmark selection, the overhead at individual nodes is determined by

the number of levels in the hierarchy. In our experiments with the anchor-based hierarchical approach, the hierarchy has 3 levels, and therefore the total number of landmarks that a node measures is upper bounded by 48. For comparison, we have experimented with 48 landmarks for the "random global" and "hybrid" landmark selection schemes. In the hybrid landmark selection, 12 out of 48 landmarks are nearby nodes. The results indicate that for the "random global" scheme, increasing the number of landmarks can slightly improve the overall prediction quality, but is not as effective as the hierarchical approach in improving the prediction accuracy for short links. The hybrid landmark selection scheme can also benefit from a larger number of landmark nodes. However, longer links still suffer from considerably degraded prediction accuracy, compared to the random global scheme and the hierarchical approach.

6. Conclusions

Recent studies suggest that the prediction quality of existing prediction mechanisms can be inadequate from the application perspective. This paper has shown that while it might be possible to improve the prediction quality through intelligent landmark selection, it is unclear how to engineer the selection procedure in order to guarantee good prediction quality over all distance ranges. Although choosing nearby nodes as landmarks can result in higher prediction accuracy for short links, longer links may suffer significantly degraded prediction accuracy.

In light of this problem, we have proposed a hierarchical approach for network distance prediction. The hierarchical prediction leverages multiple coordinates at multiple distance scales. The right scale is chosen for predicting the target distance. We study two hierarchical prediction schemes. The first scheme leverages a shared landmark hierarchy. The second scheme allows flexible landmark selection at individual nodes and the "hierarchy" is defined by increasingly smaller distance scales. Experiments with Internet measurement traces show that the hierarchical approach outperforms the hybrid landmark selection scheme: short links can be predicted with higher accuracy with little impact on the medium or long links.

References

[1] NLANR Active Measurement Project. http://amp.nlanr.net/.

[2] The P2PSim Project. http://pdos.csail.mit.edu/p2psim/.

(a) CDF of relative error (b) Distribution of relative error (c) Range accuracy (d) Reverse range accuracy
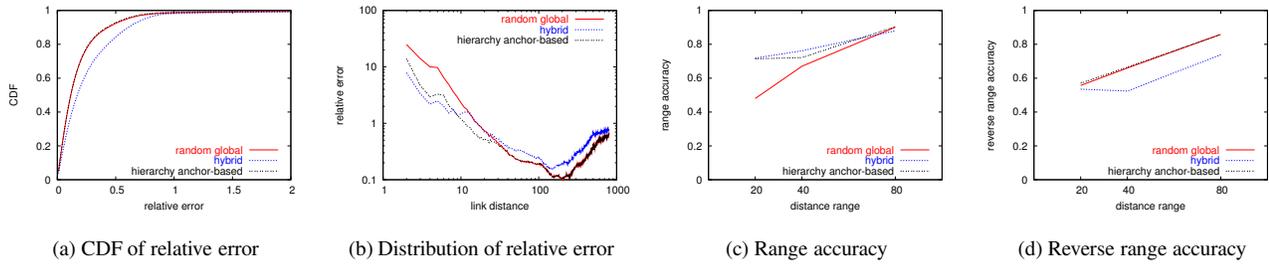
Figure 4. Prediction accuracy with the anchor-based hierarchical approach (P2PSim). The random global scheme uses the same set of landmarks as the level-0, i.e., the global level, which are randomly selected from the anchors.
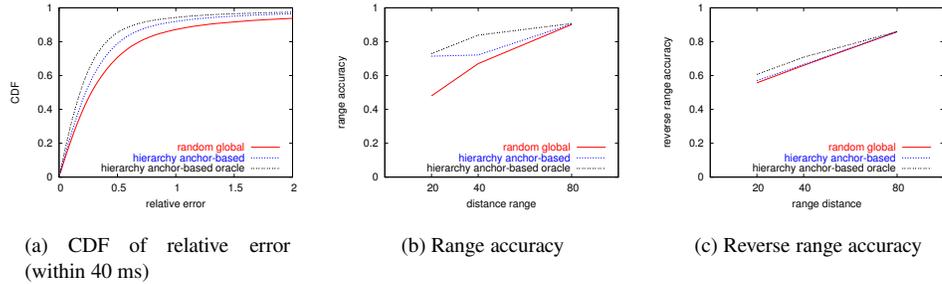

(a) CDF of relative error (within 40 ms) (b) Range accuracy (c) Reverse range accuracy

Figure 5. Prediction accuracy of the anchor-based hierarchical approach with oracle (P2PSim)

[3] Y. Chen, K. H. Lim, R. H. Katz, and C. Overton. On the Stability of Network Distance Estimation. *ACM SIGMETRICS Performance Evaluation Review*, 30, September 2002.

[4] M. Costa, M. Castro, A. Rowstron, and P. Key. PIC: Practical Internet Coordinates for Distance Estimation. In *Proceedings of IEEE ICDCS*, March 2004.

[5] F. Dabek, R. Cox, F. Kaashoek, and R. Morris. Vivaldi: A Decentralized Network Coordinate System. In *Proceedings of ACM SIGCOMM*, August 2004.

[6] P. Francis, S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavitt, and L. Zhang. IDMaps: A Global Internet Host Distance Estimation Service. *IEEE/ACM Transactions on Networking*, October 2001.

[7] K. P. Gummadi, S. Saroiu, and S. D. Gribble. King: Estimating Latency between Arbitrary Internet End Hosts. In *Proceedings of SIGCOMM Internet Measurement Workshop (IMW)*, November 2002.

[8] S. M. Holz. Routing Information Organization to Support Scalable Interdomain Routing with Heterogeneous Path Requirements, 1994. Ph.D. Thesis, University of Southern California.

[9] H. Lim, J. C. Hou, and C. ho Choi. Constructing an Internet Coordinate System Based on Delay Measurement. In *Proceedings of ACM IMC*, October 2003.

[10] E. K. Lua, T. Griffin, M. Pias, H. Zheng, and J. Crowcroft. On the Accuracy of Embeddings for Internet Coordinate Systems. In *Proceedings of ACM IMC*, October 2005.

[11] Y. Mao and L. K. Saul. Modeling Distance in Large-Scale Networks by Matrix Factorization. In *Proceedings of ACM IMC*, October 2004.

[12] T. S. E. Ng and H. Zhang. Predicting Internet Network Distance with Coordinates-Based Approaches. In *Proceedings of IEEE INFOCOM*, June 2002.

[13] T. S. E. Ng and H. Zhang. A Network Positioning System for the Internet. In *Proceedings of USENIX*, June 2004.

[14] M. Pias, J. Crowcroft, S. Wilbur, T. Harris, and S. Bhatti. Lighthouses for Scalable Distributed Location. In *Proceedings of IPTPS*, February 2003.

[15] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker. Topologically-Aware Overlay Construction and Server Selection. In *Proceedings of IEEE INFOCOM*, June 2002.

[16] Y. Shavitt and T. Tankel. Big-Bang Simulation for Embedding Network Distances in Euclidean Space. In *Proceedings of IEEE INFOCOM*, April 2003.

[17] Y. Shavitt and T. Tankel. On the Curvature of the Internet and Its Usage for Overlay Construction and Distance Estimation. In *Proceedings of IEEE INFOCOM*, March 2004.

[18] L. Tang and M. Crovella. Virtual Landmarks for the Internet. In *Proceedings of ACM IMC*, October 2003.

[19] W. Theilmann and K. Rothermel. Dynamic Distance Maps of the Internet. In *Proceedings of IEEE INFOCOM*, March 2000.

[20] M. Waldvogel and R. Rinaldi. Efficient Topology-Aware Overlay Network. In *Proceedings of ACM HotNets*, October 2002.

[21] B. Wong, A. Slivkins, and E. G. Sirer. Meridian: A Lightweight Network Location Service without Virtual Coordinates. In *Proceedings of ACM SIGCOMM*, August 2005.

[22] Z. Xu, P. Sharma, S.-J. Lee, and S. Banerjee. Netvigator: Scalable Network Proximity Estimation. Technical Report HPL-2004-28R1, Hewlett-Packard Laboratories, Palo Alto, CA, March 2005.

[23] R. Zhang, Y. C. Hu, X. Lin, and S. Fahmy. A Hierarchical Approach to Internet Distance Prediction. Technical Report TR ECE 06-03, Purdue University, March 2006. http://www.ece.purdue.edu/~ychu/publications/icdcs06_tr.pdf.

[24] R. Zhang, C. Tang, Y. C. Hu, S. Fahmy, and X. Lin. Impact of the Inaccuracy of Distance Prediction Algorithms on Internet Applications: An Analytical and Comparative Study. In *Proceedings of IEEE INFOCOM*, April 2006.

[25] H. Zheng, E. K. Lua, M. Pias, and T. Griffin. Internet Routing Policies and Round-Trip-Times. In *Proceedings of Pervasive and Active Measurement Workshop (PAM)*, March-April 2005.