PNDP: Predicting Accurate Network Distance Based on Clustering*

Li Xiao-Yong, Wang Yi-Jie, Fu Yong-Quan, Sun Wei-Dong

National Key Laboratory for Parallel and Distributed Processing, School of Computer Science, National University of Defense Technology, Changsha 410073, China

{sayingxmu, wangyijie, quanyongf, sunweidong}@nudt.edu.cn

Abstract

Network distance prediction is important in many emerging applications, such as proximity-aware content redirection in CDNs and matchmaking in network games. Network coordinate is one of the most popular way to estimate the network distance. However, existing network coordinate methods have weaknesses in accuracy and stability. According to the problem of accurate network distance predictions, we propose a novel cluster-based network distance prediction method PNDP. Firstly, it clusters the nodes with EBinning and uses optimal landmark selecting strategy. Secondly, it secures the coordinate by filtering malicious candidate landmarks. Finally, it adopts a heuristic coordinate updating mechanism and coordinate stabling mechanism to improve the convergency and stability of the coordinate computation. Experimental evaluation shows that PNDP can predict the network distance scalably and accurately, with fast convergence speed.

Keywords : network distance, network coordinate, clustering, latency estimation, landmarks selection

1 Introduction

Network distance, often measured as RTT (the Round-Trip Time), is the basis for network-aware overlay construction, proximity-aware server selection, and many other network measurement applications. However, since the large-scale and strong dynamic characteristics of distributed applications, fully measuring all-pair network distances becomes very difficult. To this end, researchers turn to estimate network distances. Network distance prediction is important in many emerging applications, such as proximity-aware content redirection in CDNs and match making in network games[1].

During last few years, network coordinate schemes have been studied extensively. However, previous proposals are still not practical enough for distributed

*This work was supported by the National Grand Fundamental Research 973 Program of China (Grant No.2011CB302601), the National Natural Science Foundation of China (Grant No.60873215), the Natural Science Foundation for Distinguished Young Scholars of Hunan Province (Grant No.S2010J5050), Specialized Research Fund for the Doctoral Program of Higher Education (Grant No.200899980003).

[1] http://halo.xbox.com/en-us

applications. Many network coordinate schemes such as GNP (Ng and Zhang, 2002) and ICS (Lim, Hou and Choi, 2003) rely on a small set of fixed infrastructure nodes, which may incur single points of failure and limit the scalability. Other proposals (Waldvogel and Rinaldi, 2002; Pias, Crowcroft and Wilbur, et al, 2003) use decentralized nodes as the landmarks to compute each nodes coordinate, which may decrease the convergency speed, and become vulnerable to falsifying coordinates of malicious peers. In addition, existing work have not fully solved the problem of coordinate drifting (Ledlie, Gardner and Seltzer, 2007), which may reduce the accuracy of distance estimation.

Accordingly, we introduce a novel cluster-based network distance prediction method PNDP. It scales well since it does not rely on any infrastructure nodes and any node with coordinates can act as a landmark. Additionally, it uses several novel coordinate optimization schemes to improve the accuracy, such as the EBinning cluster mechanism, optimal landmark selection strategy, the coordinates securing mechanism and the coordinate stabilizing mechanism. Finally, PNDP makes use of a heuristic coordinate updating mechanism to improve the convergency of the coordinate computation. Simulation results confirm that PNDP is more scalable and accurate than state-of-art coordinate schemes.

2 Predicting Network Distance with PNDP

2.1 Background

We introduce two representative network coordinates. Other network coordinate methods are similar with these methods in terms of the design principles.

GNP: GNP predicts network distances using d-dimensional Euclidean space. It computes the coordinates using the *Simplex Downhill* method (Nelder and Mead, 1965) with fixed landmarks in centralized manner. GNP does not scale well with increasing number of nodes, as it causes single points of failure. Furthermore, GNP suffers from the **TIV**(Triangular Inequality Violations) problem, which is quite common in wide-area networks.

Vivaldi: Vivaldi (Dabek, Cox and Kaashoek, et al, 2004) is fully decentralized, in that any node can be the landmark node. Each node incrementally updates its coordinate based on minimizing the estimation error of coordinate distances and RTTs with respect to other nodes. Therefore, Vivaldi scales well with increasing system size. Nevertheless, Vivaldi also suffers from the TIV problem by the Euclidean space assumption. Besides, real-world deployments show that Vivaldi is still not accurate enough (Choffnes and Bustamante, 2010).

2.2 Our Design

PNDN predicts network distances using Euclidean space for modeling, which means that PNDN suffers from the TIV problem. However, we mitigate the TIV problem through a novel technique. Additionally, PNDP uses a distributed landmark strategy to enhance the scalability and updates the coordinates to accommodate the dynamic changes of the coordinates.

We introduce Membership Server (**MS**) to store the information of nodes whose coordinates have been computed. The information containing the IP addresses of the nodes, the cluster numbers that the nodes belong to. With this centralized server *MS*, the global information can be recorded and help for the following coordinates computing process. As the amount of stored information in each node is not too large, so better performance server can easily satisfy this requirement. The distance prediction in PNDP mainly includes four processes:

1. The EBinning clustering process for the new node (*section 3.1*);
2. The landmarks selecting process for the new node (*section 3.2*);
3. The malicious landmarks filtering process (*section 3.3*);
4. The coordinates computing, updating and stabilizing process (*section 3.4*).

Firstly, the new attending node firstly visits *MS* to get some index nodes information for clustering and obtains the cluster number by using the EBinning clustering mechanism. Secondly, we adopt the optimal landmarks selecting strategy to choose some nodes in the initial set of coordinates for the subsequent calculation process. Thirdly, we secure the landmarks by filtering the malicious ones to get the final landmarks. Finally, we update and stabilize coordinates by the coordinate updating mechanism and coordinate stabilizing mechanism respectively.

3 The PNDP method

3.1 The clustering process

Clustering is widely used for network distance prediction. In this paper, we propose a novel distributed clustering mechanism EBinning. The basic idea of EBinning is that two close nodes in metric space have the similar distance to the other nodes. Similarly, if two nodes have the similar distance to the selected reference nodes, then we can believe that the nodes are close, so they can be divided into the same cluster. However, based only above idea is not enough, as it may cause false clustering problem, which means that the two nodes have the same number, but they are in different clusters. To solve this problem, EBinning uses the extended probing to avoid it, which works as follows.

1. Node A gets k index nodes from *MS*, denoted as $L_1, L_2, \ldots, L_k$;
2. Node A measures the distances $d_1, d_2, \ldots, d_k$ to the k index nodes by Ping method, where d_i denote the distance to the i-th index node;
3. Classify the distances into s levels, and the delay of i-level range in $[(i-1)*m, i*m]$, where m is a constant, and the s-level range in $[m*(s-1), \infty]$;
4. Node A determines its own level according to the k distances. For instance, given four distances to the index nodes are 123ms, 55ms, 211ms and 322ms, then the nodes initial bin number is *2134*;
5. Node A accesses a few (2-3) nodes of the cluster which corresponding to *2134* to avoid false clustering, probing the distances, and test the average distance is whether nor not smaller than δ. If so, then the bin number is the initial number, terminate; otherwise, do ahead with step 6;
6. Node A accesses to the extended $k'(k' \leq 3)$ index nodes in *MS*, and determines its own bin number as step 3 and 4, if the extended number is *ABC*, then the final bin number of A is *2134ABC*.

In the EBinning method, as for two nodes, if the final bin numbers are the same, then we can consider that the nodes are in the same cluster. The bin number here is the cluster number we discuss. After the clustering, we store the cluster number to *MS* to facilitate the following processes.

3.2 Landmarks selecting strategy

PNDP employs a distributed landmarks that any nodes whose coordinates have been computed can be as landmarks for other nodes. Consequently, it greatly improves the scalability. To further examine the effect of different landmarks selecting strategies, we have implemented and compared three strategies to pick landmarks, which are described as follows and shown in Figure **1**.

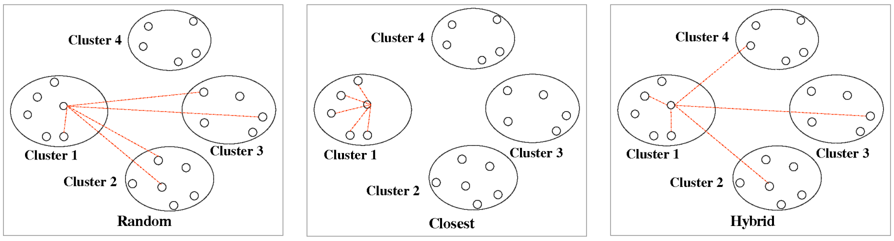

Figure 1 Different landmark selection strategies

- *Random*: randomly choose k nodes as landmarks with uniform probability;
- *Closest*: pick the k closest nodes in the cluster as the landmarks;
- *Hybrid*: pick some nodes as in *Random* and others as in *Closest*.

The experimental results will be discussed later in section 4.1. Based on the comparing, PNDP uses the best landmarks selecting strategy to improve the accuracy of the coordinate computation. These strategies are similar to the three landmark selecting strategies in PIC (Costa, Castro and Rowstron, et al, 2004). However, due to PNDP with novel clustering mechanism, the strategies are essentially different. As for the *Closest* strategy, if the number of nodes in one cluster is small than k, then we can select landmarks in the nearby cluster. As for the *Hybrid* strategy, we choice the landmarks distributed in various clusters as possible, making the choice presented a better distribution in the global. The experimental results in section 4.1 confirm that the *Hybrid* strategy achieves lower relative errors than the other strategies.

3.3 Secure the landmarks

Network coordinate predicting methods are vulnerable to malicious nodes (Ledlie, Gardner and Seltzer, 2007). Once the malicious nodes are selected as landmarks, they can lie about their coordinates or interfere with the distance measurement. The results of these attacks are a set of coordinates that can be arbitrarily wrong. Thus, we introduce a novel mechanism to detect malicious nodes for improving the landmarks before calculating the coordinates, which enhances the network coordinate security and makes the results more accurate. Assume that we know the candidate index node set $L = \{L_1, L_2, \ldots, L_n\}$, the coordinates of the index nodes C_i, and relative error ratio threshold ε , through the process. The malicious nodes filtering process is shown as follows.

1. Choose $d+1$ index nodes from L to attend the computing and work out the coordinate C_H for the new attending node H;
2. Compute prediction distances $p_{HL_i}, (1 \le i \le n)$ to $L_1, L_2, \ldots, L_n$ for node H, where $p_{HL_i} = \| C_H - C_{L_i} \|$;

3. Compute the relative error of the prediction distance between H and L_i , which is $relerr_{HL_i} = \frac{\|p_{HL_i} - d_{HL_i}\|}{d_{HL_i}}$;
4. Compute the mean relative error of the prediction distance between H and all other nodes, which is $relerr_{avg} = \frac{\sum_{i=1}^{n} relerr_{HL_i}}{n}$;
5. If the relative error ratio of H for any index node i satisfies $\mu = \frac{relerr_i}{relerr_{avg}} > \varepsilon$, then i is considered as a malicious node and we eliminate it from L;
6. Use the above method to eliminate the malicious nodes in L, until the number of the index node meets $n \neq d + 1$.

3.4 Dynamic Coordinates computation

3.4.1 Coordinates updating

PNDP computes the initial coordinates of all the nodes using *Simplex Downhill* method. To meet dynamic changes in the environment, we propose a novel heuristic for updating network coordinates. Unlike Vivaldi, nodes in PNDP firstly interact with some landmarks and put different error weight factors to the landmarks according to the standard errors to them. Secondly, nodes use the standard errors with various weighting factors to recalculate the total weighted error. Thirdly, nodes adopt *Simplex Downhill* method to minimize the total error value and compute the corresponding coordinates. Finally, nodes employ the verification process to check for the new total error with the new coordinates and adaptively updates the coordinates.

The specific description of the coordinate updating process as follows.

1. Inquiry all the index nodes and compute the **relative error** $rerr_j = \frac{|d_{ij} - \|C_i - C_j\||}{min(C_i, C_j)}$, where d_{ij} is probing distance of i and j, C_i and C_j denote initial coordinates of the current node and the index node respectively.
2. Compute **error weight** w_j of each index node and the **total prediction error** E_i, where $w_j = \frac{rerr_j}{\sum_{k \in samples} rerr_k}$, $E_i = \sum_{k \in samples} |d_{ik} - \|C_i - C_j\||^2$
3. Compute the new coordinate C_i', which minimizes the weighted total prediction error $\sum_{k \in samples} w_k \left(d_{ik} - \|C_i' - C_k\|\right)^2$
4. Check for the new prediction error $E_i' = \sum_{k \in samples} |d_{ik} - \|C_i' - C_j\||$
5. According to the relationship of E_i' and E_i to adjust the coordinate $C_i^{new} = C_i + \varepsilon(C_i' - C_i)$, where $\varepsilon = min(\frac{E_i'}{E_i}, 1)$

3.4.2 Coordinates stabilizing

Ledlie, Gardner and Seltzer(2007) have found that network coordinates are drifting. It means that the coordinates will gradually deviate from the original position after a long period, which will result in the coordinatesrealignment.

We stabilize the coordinates using the clustering again, since the cluster structure keeps relative stable due to the stationary of delays. After a relatively long period of time, PNDP runs the coordinates stabilizing process to stabilize the network coordinates to prevent the occurrence of coordinates drifting. Assume that we know all the nodes coordinates in each Binning, the related probing distances, and the given average relative error threshold σ.The specific coordinates stabilizing process is shown as follows.

1. Compute the average coordinate $\bar{c}$ of all the coordinates in each cluster;

2. Choose the closest node to $\bar{c}$ in each cluster as the reference node for stabilizing, labeled as l_i ;
3. Detect all the clusters by the above way and obtain a set of reference points $L = l_1, l_2, \ldots, l_k$;
4. For each node in L, compute the relative error $relerr_j$ to other nodes every period T, and the average relative error $relerr_{avg}$ to all the other nodes;
5. if $relerr_{avg} > \sigma$, recompute the coordinate according to the other nodes in L and update the coordinates for all nodes in the cluster;
6. Use the same way to detect all the nodes in L and the corresponding coordinates updating, until fininsh detecting all the ndoes in L.

4 Experimental evaluation

In the experimental evaluation, we use the MIT King data set consisting of pairwise delays between 1740 DNS servers, which are collected by the King method (Gummadi, Saroiu and Gribble, 2002) for our simulation study. We compare PNDP with GNP and Vivaldi methods and use three metrics for comparison: the traditional metric Ralative Error and two new metrics SRRL (Smart Relative Rank Loss) and ECNL (Extended Closest Neighbors Loss), which are proposed by Key, Thomson and Thomson (2008). SRRL is used to describe the loss distance order, and it to some extent relaxes the loss of the order. ECNL to some extent tolerates the error for the closest nodes by prediction.

4.1 Landmark strategies evaluation

We evaluate three landmark selecting strategies in PNDP. As shown in Figure **2**, we can see that the accuracy of the *Random* and *Closest* are close, but *Hybrid* is best among all strategies. It implies that *Random* landmark having the best distribution in the overall, while the *Closest* strategy is to focus on the landmark choice of locality. Both the two landmark strategies are two extremes, could easily lead to inconsistencies in the coordinates. Since *Hybrid* combines the advantages of *Random* and *Closest*, it achieves much lower errors.

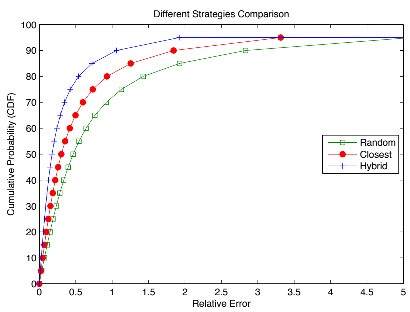

Figure 2 Comparison of different landmark strategies

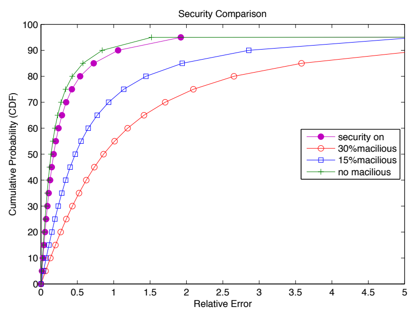

Figure 3 Comparison of coordinate security

4.2 Security Evaluation

We have tested the security of PNDP. For testing the safety of coordinates, we introduce malicious nodes in different proportions to study the influence of malicious nodes in the performance of coordinates. While in the test of convergency of coordinates, we compare PNDP and Vivaldi with the accuracy under different updating times with $d = 2$, $C_c = 0.25$. Figure **3** shows that malicious nodes that have detected by PNDP, its performance are very close to the performance with non-malicious. Compared to GNP and Vivaldi in terms of security, the security of PNDP has significantly improved.

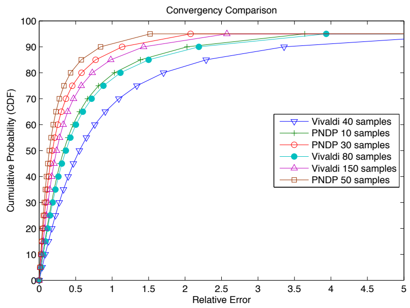

Figure 4 Convergency Test

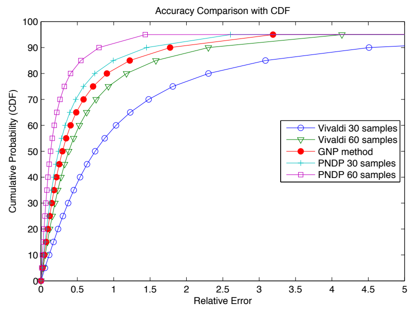

Figure 5 Accuracy Comparison

4.3 Convergency Comparison

Next we test the convergency of PNDP. We also plot the convergence of Vivaldi for comparison. From Figure **4**, we know that the performance of PNDP which after sampling 50 times is far exceed Vivaldi after sampling 150 times. Compared to the classic method Vivaldi, the convergence speed and accuracy of PNDP has obvious advantages.

4.4 Accuracy Evaluation

Figure **5** shows that PNDP is the most accurate one under the metric of relative error. After updating 60 times, the relative error below 0.5 for PNDP accounts for nearly 90%. Compared with other methods, its accuracy is higher. Meanwhile, Figure **6** shows that PNDP achieves the best SRRL (d_{min} is set to 10ms) after 50 times sampling. Figure **7** show that PNDP has much lower ECNL ($\varepsilon = 5\%$) value after 50 times. From all the comparisons above, we can confirm that PNDP has the highest accuracy.

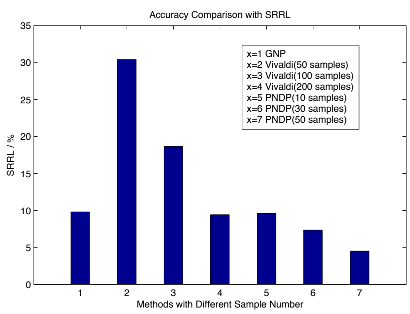

Figure 6 Accuracy with SRRL

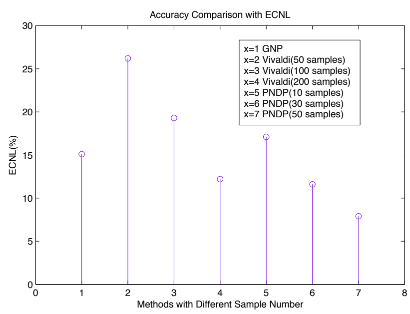

Figure 7 Accuracy with ECNL

5 Conclusion

This paper presents PNDP, a novel network coordinate method for estimating the network distances in a distributed manner. PNDP is scalable as it adopts distributed landmarks and does not rely on fixed infrastructure nodes. PNDP is also robust as it can secure the coordinates by filtering the malicious nodes and it can compute accurate coordinates even when part of nodes are malicious. PNDP improves the accuracy and convergency by a coordinate stabilizing mechanism and a heuristic coordinate updating mechanism. We have verified that the performance of PNDP is better than the classical method GNP and Vivaldi with the relative error, SRRL and ECNL metrics. Therefore, PNDP is more practical for many large-scale distributed applications like network-aware overlay construction and resources location in the network.

References

[1] Choffnes D. R. and Bustamante F. E., Pitfalls for Testbed evaluations of internet system, *Sigcomm Compu. Commu. Rev.*, 40(2010), 43-50.

[2] Costa M., Castro M., Rowstron A. and Key P., PIC: practical Internet coordinates for distance estimation, *Proc. ICDCS*, 24(2004), 178-187.

[3] Dabek F., Cox R., Kaashoek F. and Morris R., Vivaldi: A decentralized network coordinate system, *Sigcomm Compu. Commu. Rev.*, 4(2004), 15-26.

[4] Gummadi K. P., Saroiu S., and Gribble S. D., King: Estimating latency between arbitrary internet end hosts, *Proc. ACM Sigcomm Workshop on Internet measurment*, 2002, 5-18.

[5] Key P., Thomson L. M. and Thomson D. T., Non-Metric coordinates for predicting network proximity, *Proc. IEEE Infocom*, 2008, 1840-1848.

[6] Lim H., Hou J. C. and Choi C. -H. Constructing internet coordinate system based on delay measurement, *Trans. Network.*, 13(2005), 513-525.

[7] Ledlie J., Gardner P. and Seltzer M., Network Coordinates in the Wild, *Proc. NSDI*, 2007.

[8] Nelder J.A. and Mead R., A simplex method for function minimization, *Jour. Compu.*, 7(1965), 308-313.

[9] Ng T. S. E. and Zhang H., Predicting Internet network distance with coordinates-based approaches, *Proc. IEEE Infocom*, 1(2002), 170-179.

[10] Pias M., Crowcroft J., Wilbur S., Harris T. and Bhatti S., Lighthouses for scalable distributed location, *Lect. notes. Compu. Sci.*, 2003, 278-291.

[11] Waldvogel M. and Rinaldi R., Efficient Topology-Aware Overlay Network, *Compu. Commun. Rev.*, 33(2003), 101-106.