Global Network Positioning: A New Approach to Network Distance Prediction

T. S. Eugene Ng and Hui Zhang, Department of Computer Science, Carnegie Mellon University*

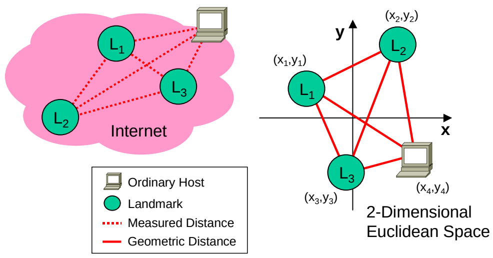

Figure 1: Computing GNP coordinates.

We propose a new approach to the network distance (round-trip transmission and propagation delay) prediction problem called Global Network Positioning (GNP). We demonstrate that it is feasible to inexpensively model the Internet as a geometric space (e.g. a 3-dimensional Euclidean space), and characterize the position of any host in the Internet by a point in this space such that the network distance between two hosts can be predicted accurately by the geometric distance implied by their coordinates in this space.

Figure 1 shows how GNP efficiently maps Internet hosts to points in a geometric space. The key technique is to first compute the coordinates of a small distributed set of cooperating hosts called Landmarks based on measured inter-Landmark distances such that the overall error between the measured distances and the geometric distances is minimized. The Landmarks' coordinates are disseminated to ordinary hosts and serve as a frame of reference. With this frame of reference, each ordinary host measures its distances to the Landmarks (the Landmarks are passive) and computes a set of coordinates for itself that minimizes the overall error between the measured and the geometric host-to-Landmark distances. The computations of Landmark and ordinary host coordinates can be cast as generic multi-dimensional global minimization problems that can be approximately solved by many available methods.

Although off-line pre-computations are required to derive the coordinates of Landmarks and hosts, modeling the Internet as a geometric space and communicating distance information using coordinates have several advantages over the traditional approach of modeling the Internet as a simplified topology and communicating distance information using individual path distances. First of all, a distance prediction in the geometric space model is simply an evaluation of the distance function which is generally both straight-forward to implement and extremely fast to compute comparing to a shortest path search in the topology model. Secondly, in a multi-party application, the distances of all paths between K hosts can be efficiently communicated by K sets of coordinates of size D each (i.e. $O(K \cdot D)$ of data), where D is the dimensionality of the geometric space, as opposed to $K(K-1)/2$ individual distances (i.e., $O(K^2)$ of data). Thirdly, host coordinates are relatively fixed local properties that can be exchanged easily among hosts when they discover each other, allowing network distance predictions to be locally computed by end hosts in a timely fashion. Fourthly, Landmarks are very simple and non-intrusive, hence they are easy to deploy and ordinary hosts behind firewalls can also participate. Finally, we can exploit the structured nature of the coordinates to efficiently perform interesting operations such as nearest neighbors searches.

*Email addresses: {eugeneng, hzhang}@cs.cmu.edu

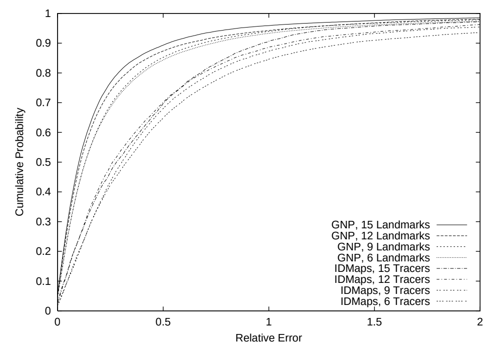

Figure 2: Relative error of GNP and IDMaps.

To evaluate GNP, we apply it to off-line collected Internet distance measurements. In the last week of May 2001, we measured the distances between 19 distributed probes and the distances between each probe and 869 Internet hosts distributed over the globe. We then conduct GNP experiments over this data by clustering the 19 probes and selecting a subset of them as Landmarks, and using the remaining probes and the 869 Internet hosts as ordinary hosts. To measure how well the predicted distance between two ordinary hosts matches the corresponding measured distance, we use a metric called relative error that is defined as $\left|\frac{predicted\ distance - measured\ distance}{\min(measured\ distance, predicted\ distance)}\right|$.

We experimented with the 5-dimensional Euclidean space model and varied the number of Landmarks from 6 to 15. Each experiment was repeated multiple times with slightly different Landmarks and the overall result is reported. For comparison, we also evaluated the performance of the current state-of-the-art distance prediction approach IDMaps [1] when applied to our data, using the corresponding Landmark nodes as the IDMaps Tracers. Figure 2 shows the cumulative probability distribution functions of the relative error for GNP (top 4 lines) and IDMaps. As can be seen, GNP predicts distances more accurately than IDMaps, and benefits consistently from the addition of Landmarks. Using GNP with 15 Landmarks, 90% of all distance predictions are within a relative error of 0.53.

GNP makes it possible to provide scalable, fast, and effective network performance optimization in distributed network services and applications because it requires no on-demand network measurements. We are working towards improving the GNP techniques and understanding how underlying Internet properties affect GNP's performance.

References

[1] P. Francis, S. Jamin, V. Paxson, L. Zhang, D.F. Gryniewicz, and Y. Jin. An architecture for a global Internet host distance estimation service. In *Proceedings of IEEE INFOCOM '99*, New York, NY, March 1999.