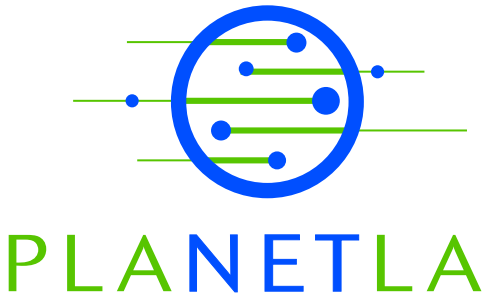


The Interdomain Connectivity of PlanetLab Nodes

Suman Banerjee
University of Wisconsin, Madison

Timothy G. Griffin
Intel Research, Cambridge UK

Marcelo Pias
Intel Research, Cambridge UK

PDN–04–019
February 2004

Appears in: *Proceedings of the Passive & Active Measurement Workshop* (PAM2004), Antibes Juan-les-Pins, France, April 2004.

The Interdomain Connectivity of PlanetLab Nodes

Suman Banerjee*, Timothy G. Griffin**, Marcelo Pias***

Abstract. In this paper we investigate the interdomain connectivity of PlanetLab nodes. We note that about 85 percent of the hosts are located within what we call the *Global Research and Educational Network* (GREN) — an interconnected network of high speed research networks such as Internet2 in the USA and Dante in Europe. Since traffic with source and destination on the GREN is very likely to be transited solely by the GREN, this means that over 70 percent of the end-to-end measurements between PlanetLab node pairs represent measurements of GREN characteristics. We suggest that it may be possible to systematically choose the placement of new nodes so that as the PlanetLab platform grows it becomes a closer and closer approximation to the Global Internet.

1 PlanetLab

The primary goal of PlanetLab is to provide a geographically distributed platform for overlay-based services and applications [12, 1]. Currently there are over 120 participating sites with a total of over 300 hosted nodes. The open and shared nature of PlanetLab is enabling an exciting range of innovative experimentation in overlay techniques. Without its collectively supported infrastructure most participating research groups would not have the means to set up such a rich environment.

For many of the same reasons, the PlanetLab infrastructure has attracted researchers working with measurements of the *legacy Internet*. That is, PlanetLab nodes are employed to actively probe and collect various Internet metrics. We argue that this is an *opportunistic* use of PlanetLab in the sense that providing an Internet measurement infrastructure has never been among the primary goals of the project. This does not mean that PlanetLab is not a useful Internet measurement platform, only that PlanetLab measurements cannot *automatically* be taken as representative of the global Internet.

In this paper we investigate the interdomain connectivity of PlanetLab nodes. We note that about 85 percent of the hosts are located within what we call the *Global Research and Educational Network* (GREN) — an interconnected network of high speed research networks such as Internet2 in the USA and Dante in Europe. Since traffic with source and destination on the GREN is very likely to be transited solely by the GREN, this means that over 70 percent of the end-to-end measurements between PlanetLab node pairs represent measurements of GREN characteristics. Whether or not such measurements are representative of the global Internet is something that needs more investigation.

This should in no way be misconstrued as a criticism of PlanetLab — we are only stating that those using PlanetLab for measurements of the global legacy Internet need

* University of Wisconsin, Madison. suman@cs.wisc.edu
** Intel Research, Cambridge UK. tim.griffin@intel.com
*** Intel Research, Cambridge UK. marcelo.pias@intel.com

to present their arguments with care. On the other hand, the GREN is in some respects more attractive to measurement researchers than the Internet at large. Primarily this is because it is more *transparent* — that is, there is more publicly available information about the connectivity of the GREN and more willingness on the part of its operators to share information with the research community. We suggest that it may be possible to systematically choose the placement of new nodes so that as the PlanetLab platform grows it becomes a closer and closer approximation to the Global Internet. We advocate that there is still a large amount of untapped diversity within GREN which can be explored for similar effect on the PlanetLab.

We present some preliminary results on a case study. We generate a site-to-site distance matrix for PlanetLab sites where distances between sites is taken to be minimum round trip time. We then enumerate all possible triangles formed by three sites and investigate the violations of the triangle inequality. Low violations are important for the feasibility of various proposals to generate synthetic coordinate systems for the Internet based on round trip time measurements [11, 13, 17]. We find that when we classify triangles as "research triangles" (all nodes on the GREN), "commercial triangles" (all nodes off of the GREN), and "mixed triangles" (combination of commercial and GREN sites). We find that the distribution of "bad triangles" is lowest for research triangles (about 12 percent), higher for mixed triangles (about 20 percent), and highest for commercial triangles (about 25 percent).

2 The Global Research and Education Network (GREN)

The global Internet is comprised of a large collection of autonomously administered networks. Some of these networks are operated by commercial enterprises, while others are operated by nonprofit organizations. Perhaps the largest nonprofit networking organizations exist to provide connectivity between research and academic institutions. This includes the Geant backbone network run by the Dante organization in Europe and the Abilene backbone network run by Internet2 in North America. Such backbones connect many regional and national research networks into a large global network providing connectivity between diverse academic and research organizations. We refer to this network as the *Global Research and Education Network* (GREN).

Figure 1 presents a simplified picture of the current GREN. The GREN is *not* by any means a single administrative entity. All of the networks are independently administered, and exhibit various degrees of cooperation. There is also large diversity in the primary goals of these networks. Some regional networks are targeted toward a specific set of users, while others serve a larger research and education community. For example, the CERN network exists largely to provide high bandwidth connectivity between physics laboratories around the world. On the other hand, the WiscNet of Wisconsin (http://www.wiscnet.net) exists to provide connectivity between K-12 educational and university level institutions. Some GREN networks provide transit to commercial network providers, while others do not. For example, the backbone of Internet2, the Abilene network, does not provide commercial connectivity, while WiscNet does. The ARENA project (http://arena.internet2.edu) provides a very useful online compendium of information about the networks making up the GREN. It should also be remembered

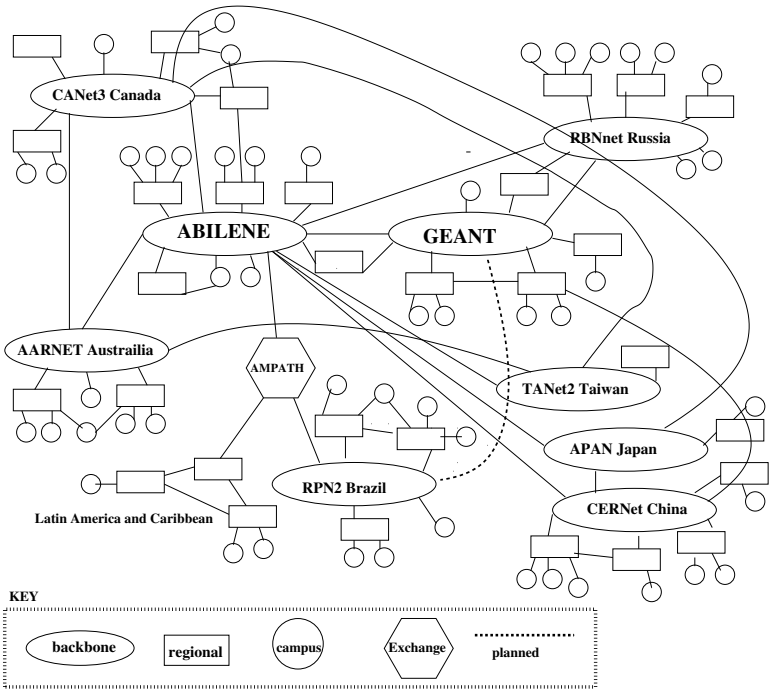


Fig. 1. A highly simplified and incomplete schematic of the GREN.

that the GREN is continually changing and expanding. For example, to Latin America from Europe, GREN traffic is routed today through the USA (Florida). A directed link between Geant and the Latin American research networks is planned within the scope of the ALICE project [15]. This will create a "short-cut" between these networks, which is likely to have an effect on how experimental results should be interpreted. Similarly, other new links are planned between Europe and USA.

The logical connectivity of the GREN networks does not immediately tell us how traffic between them is *routed*. Figure 2 depicts two sites A and B that have connectivity to both the GREN and the commercial Internet. One would normally expect that both sites A and B prefer routes from the research network over routes learned from their commercial providers. There are several reasons for this, the primary one being that research connectivity is normally paid for out of a special source of funds, and it may not be metered as it is normally done with commercial traffic. In addition, there may be the perception that for research work, the research network will give better performance. In this way, we can think of the research network as providing a "short cut" between A and B that bypasses the commercial Internet. Of course, this may be occasionally overridden by local routing policies. The actual connectivity of most sites is more complex than Figure 2. For example, Figure 3 shows the connectivity of two PlanetLab sites: planetlab2.cs.unibo.it and planetlab2.cs.wisc.edu. We have verified our assumptions about routing policies with "AS level traceroutes". For the above example this yields:

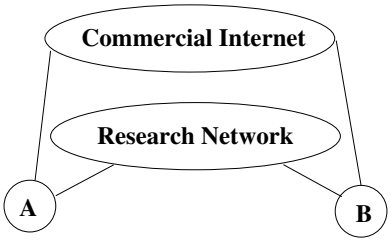


Fig. 2. Simple schematic of the routing choices as A and B

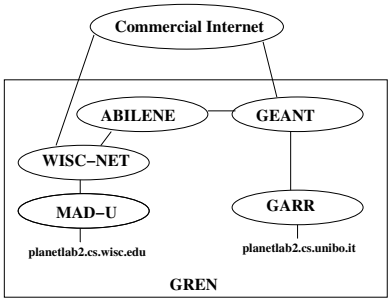


Fig. 3. A slightly more complicated example.

```
AS 137   : GARR    --- Italian REN                (5 routers)
AS 20965 : GEANT   --- Dante Backbone             (3 routers)
AS 11537 : ABILENE --- Internet2 backbone         (2 routers)
AS 2381  : WISCNET --- Wisconsin REN              (3 routers)
AS 59    : MAD U   --- U. of Wisconsin, Madison   (5 routes)
```

We compute these by first performing router-level traceroutes between PlanetLab nodes using Scriptroute [2] and then post-processing the results to associate intermediate routers with ASNs. With only a few exceptions, our expectations about routing were confirmed.

3 PlanetLab and the GREN

Since those engaged in the building of PlanetLab and the deployment of overlay services are largely academically oriented researchers, it should not be surprising that the majority of PlanetLab sites are located in the GREN. However, as far as we know there has never been an attempt to systematically quantify this relationship.

We must determine whether a given PlanetLab host or site is connected to the GREN or the commercial Internet. We use the following approach. First, we obtain a reasonable list of ASes in the GREN. This was done by extracting from Route-Views [3] BGP routes only those routes that were announced by Abilene (this can be found in the "oix full snapshot" source). This BGP session with Route-Views announces only research destinations (about 8,000 routes). In a perfect world, the ASNs in the AS-PATHS of these routes would be exactly the ASNs of the GREN. However, a small number of routes from the commercial Internet get "leaked" into these tables (for reasons still unclear), and so we must eliminate some ASes that are known to be commercial providers (Sprint, MCI, and so on). On January 15, 2004, this procedure left us with a list of 1,123 ASNs for the GREN.

Next, for each IP address associated with a PlanetLab node, we find the originating ASN associated with the longest matching prefix. In order to generate a reasonable mapping of originating ASNs to prefixes, we merged routing information from 20 BGP tables taken from RIPE [10] and Route-Views. This is essentially the basis of the technique described in [9, 8] for obtaining AS-level traceroutes. If the originating AS is in

our GREN list, we classify the node as a GREN node. Otherwise it is a commercial node. If the address is associated with Multiple Originating ASNS, some being commercial and others research, we classify the site as MOAS. With the 276 production nodes of January 15, 2004, we obtain the following breakdown as shown in Table 1.

class	number	percent
MOAS	14	5.1
Commercial	27	9.8
Research	235	85.1

Table 1. Breakdown of hosts.

class	number	percent
MOAS	7	5.7
Commercial	12	9.8
Research	104	84.5

Table 2. Breakdown of sites.

We then group the hosts into *sites* containing multiple hosts. This produces 123 sites with a breakdown very close to that of the hosts (Table 2).

Note that this means that over 70 percent of the end-to-end measurements between PlanetLab node pairs represent measurements of GREN characteristics. We then classified these sites into one of four broad geographical regions: North America (NA), Europe the Middle East and Africa (EMEA), Asia Pacific (AP), and Latin America (LA). The breakdown of the sites are listed in Table 3.

region	number	percent
LA	1	0.8
AP	6	4.9
EMEA	25	20.3
NA	91	74.0

Table 3. Breakdown by region.

class	region	number	percent
Commercial	EMEA	1	0.8
Research	LA	1	0.8
Research	AP	6	4.9
Commercial	NA	11	8.9
Research	EMEA	23	18.7
Research	NA	74	60.2
Research	NA or EMEA	97	78.9

Table 4. Combination of regions and types.

We now look at the combinations of these two classifications (Table 4). In this table we have ignored the MOAS sites (although they are still counted when computing percentages) Note that this means that over 60 percent of the end-to-end measurements between PlanetLab node pairs represent measurements of the NA and EMEA portion of the GREN. In summary, we note that not only are PlanetLab nodes situated in the GREN corner of the Internet, but inhabits a fairly small part of that world as well. The majority of site-to-site data flows are carried over the Abilene and Geant networks. Contrast this with the rich diversity illustrated in Figure 1.

3.1 Case Study — The Triangle Inequality

A number of applications would benefit from a global topological map of the Internet, which incorporates information such as the number of hops and the network latency between Internet hosts. Towards this goal, research in recent years have focused on distributed techniques to construct an *Internet Coordinate System,* in which each host can be assigned a virtual coordinate and the distance between the virtual coordinates of any two hosts will approximate their actual distance for some specific network performance

metric. Network latency is an example of such a metric. There may be one such coordinate system for each metric of interest. Some of the proposed techniques are Global Network Positioning (GNP) [11], Lighthouses [13], Virtual Landmarks [17], ICS [7] and Practical Internet Coordinates (PIC) [4]. The goal in these techniques is to significantly reduce the number of measurements required, thereby scaling to large number of hosts. There are many applications that can leverage such an Internet Coordinate System, e.g. topologically-aware construction of peer-to-peer systems [5], network latency estimation tools [16], and distributed resource discovery mechanisms [14]. The general problem to construct such a coordinate system can be abstracted as the problem of mapping, or *embedding*, a graph into a metric space. A metric space M is defined by the pair (X, d) where X represents the set of valid objects and d is a metric. A metric is a function $d : X \times X \rightarrow \mathbb{R}$ such that for $x_i, x_j, x_k \in X$, d satisfies the following properties:

1. $d(x_i, x_j) \geq 0$ (positiveness),
2. $d(x_i, x_j) = d(x_j, x_i)$ (symmetry),
3. $d(x_i, x_j) \leq d(x_i, x_k) + d(x_k, x_j)$ (triangle inequality)

In contrast, a vector space V is a particular metric space in which $X = \mathbb{R}^k$ and it has distance function D that satisfies properties (1), (2) and (3). The embedding problem consists of finding a scalable mapping $\gamma : X \rightarrow \mathbb{R}^k$ that transforms *objects* $\{x_1, ..., x_n\}$, of the original space, in this case the Internet graph, onto *points* $\{v_1, ..., v_n\}$ in a target vector space V of dimensionality k.

The concept of distortion is used to measure the quality of an embedding technique. It measures how much larger or smaller the distances in the vector space $D(v_i, v_j)$ are when compared to the corresponding distances $d(x_i, x_j)$ of the original space. The distortion is the lowest value $c_1\, c_2$ that guarantees that [6]:

$$\frac{1}{c_1} \,.\, d(x_i, x_j) \leq D(v_i, v_j) \leq c_2 \,.\, d(x_i, x_j)$$

Different techniques have been recently proposed to embed the Internet graph into a metric space, while trying to minimize the distortion at the same time [11, 13, 17, 7, 4, 5]. The distance metric used in these techniques is the minimum round trip time (RTT) over a significantly large time interval.

3.2 Triangle Inequality on the Internet

The AS-level traffic on the Internet is forwarded based on dynamic BGP routing policies. In general, service providers are free to set their own BGP policies and make local arrangements with peering providers and customers. Moreover, service providers are often multi-homed, which means they have multiple connections to the Internet for various reasons such as links resilience by using backup links and traffic load balancing. Also, the customer-provider relationship can be regarded as one reason why shortest path routing may not be used in a given fraction of the network. There will be cases where providers, for business-related reasons, will prefer to route traffic via another provider as opposed to a customer, even if the shortest path is through its customer.

Thus, there is no reason to expect that the triangle inequality, an important property of metric spaces, holds on the Internet. In this section we enumerate all possible triangles formed by three PlanetLab sites. We then analyze the violations of the triangle inequality, as well as a measure of how good these triangles are.

3.3 Data and Methodology

We used RTT data[1] collected between all PlanetLab hosts from November 17th to November 23rd 2003. We then calculated the minimum RTT between each pair of hosts available between those dates on consecutive 15-minute periods. Thus for each day in this period there were 96 matrices of RTT measurements (each entry represented a pairwise RTT measurement), and the size of each matrix was 261×261. Over the seven day period we therefore had 672 such matrices.

Our goal in this evaluation was to identify the various triangle inequality violations between these pairs of hosts. Since we wanted to perform this computation over our estimate of propagation delays on the paths only (and not the queueing and congestion effects), we first constructed a single matrix, in which each entry represented the minimum RTT between a pair of PlanetLab nodes over the entire 7-day period. This avoided biases in the results due to high variations in RTTs, e.g. during congested periods. Our analysis of the data indicated that by taking the minimum over a 7-day period, we can filter out congestion related effects which have periodic weekly patterns.

Many PlanetLab sites had multiple nodes per site. For instance, the Computer Laboratory (University of Cambridge) site hosted three nodes `planetlab1`, `planetlab2` and `planetlab3` under the domain `cl.cam.ac.uk`. The minimum RTTs between nodes within a site were very small, often of the order of 1 ms.

Examining triangles between all triples of nodes was therefore wasteful and biased our results by the distribution of nodes in PlanetLab sites. Thus, we selected a representative node in each site so as to build a *site by site* matrix D', reducing the distance matrix to 123×123 RTTs.

3.4 Preliminary Results

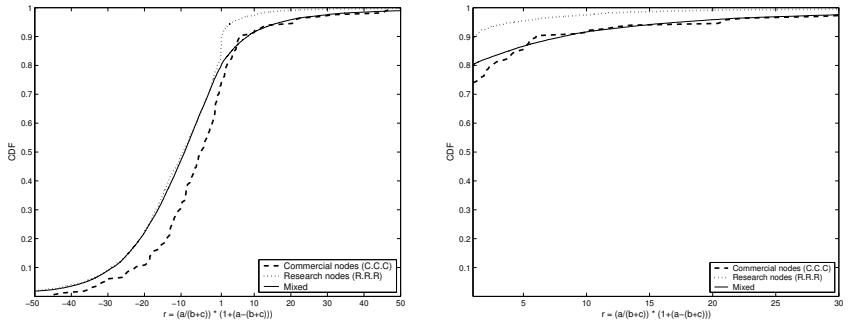
We defined a metric r, termed the *goodness* of a triangle, to test violations of the triangle inequality on the D' matrix. By convention a is the longest side of a triangle, and the other remaining sides are termed b, c. The metric r is defined as:

$$r = \left(\frac{a}{b+c}\right) \times (1 + (a - (b+c)))$$

This metric is made of two terms. The first one distinguishes between 'good' and 'bad' triangles (violators). Values of r less or equal to 1 represent 'good' triangles. In contrast, any triangle with r greater than 1 is a *violator*, i.e. the relation between its sides lengths do not satisfy the triangle inequality property. This first term is multiplied

[1] The all-pairs RTT measurement data was being continuously collected by Jeremy Stribling (MIT) and it is available from `http://www.pdos.lcs.mit.edu/~strib/pl_app/`

by a factor which tells the 'goodness' of a triangle. The higher the value of r, the worse it is the triangle. We believe that triangles with higher r have higher impact on the embedding distortion factors c_1 and c_2. However, further research is required to quantify such a distortion.



(a) Distribution of all triangles (b) Distribution of violators

Fig. 4. CDF of the goodness metric r

In Figure 4, we plot the CDFs of metric (r) of triangles derived from the upper triangular part of D'. The three lines in each graph are: (1) RRR, which includes triangles formed only by sites in GREN, (2) CCC, which includes only those triangles that are formed by sites in the commercial Internet, and (3) Mixed, which includes all the other triangles (combination of commercial and GREN sites). Figure 3.4 shows the distributions of 'good' and 'bad' triangles; whereas Figure 3.4 presents only the cases of violators, i.e. values of r greater than 1.

We can make two interesting observations from this figure. First, there are fewer violations of the triangle inequality in the GREN than in the commercial Internet. Violators represent about 12% of the cases in the GREN and about 25% for the commercial Internet. When mixed triangles were tested, around 20% of them were violators.

And second, the triangles in the GREN are 'better' formed than the triangles in the commercial Internet. This means that 'bad' triangles in the commercial Internet tended to be larger than the ones of the GREN and also have a very long side.

At this stage, we speculate that this difference might be the reason why the distribution of violators for mixed triangles is closer to the distribution of commercial triangles (Figure 3.4). One would expect the opposite, as the number of GREN sites forming mixed triangles are predominantly larger.

It seems that commercial sites when combined to GREN sites influenced the shape of the triangles, i.e. distorting them. The nature and cause of these differences need further detailed investigation.

4 Recommendations for improving PlanetLab diversity

The results presented in our case study leads to broad research questions — How can we choose a set of nodes in the Internet so that inter-node experiments can reasonably

be taken to represent behavior on the Internet as a whole? In the PlanetLab context this might be reformulated as follows — How can we systematically choose the placement of new nodes so that as the PlanetLab platform grows it becomes a closer and closer approximation to the Global Internet?

One aspect of this is the distinction between commercial and research networks. One direction towards diversity might be to increase the number of sites on the commercial networks. This may be desirable, but as a general strategy it has several shortcomings. The main problem with this approach is that many research sites are buried deep inside of department/campus/regional networks and find it difficult if not impossible to get direct access to the commercial Internet. For example, the PlanetLab hosts at the computer science department in UW Madison are many router hops away from a commercial router. It obtains connectivity through a departmental network, then a campus network, and finally through a state-wide research network (WiscNet). It is only WiscNet that provides access to the commercial Internet. A solution to add diversity in PlanetLab paths might be to work with the upstream networks and have some PlanetLab net blocks announced only to the commercial networks upstream from Wiscnet. Perhaps this could be implemented with some type of BGP community sent up on the routes when they are announced into WiscNet. It probably can be done, but requires a lot of BGP magic.

We would advocate a different approach. The entire GREN is very large and comprises of a diverse set of networks. For example, CERN runs a network that is to GREN what the GREN is to the global Internet. Out of the 1123 ASes that we found in the GREN, PlanetLab nodes are located in 80, i.e. less than 8% of them. Perhaps the best strategy (in terms of cost and feasibility) is to make a targeted and systematic attempt to exploit the existing diversity of the GREN. Our exploration and research of the characteristics of GREN lead us to believe that there are a lot of diverse networks within GREN itself. For example, While some research divisions in academic environments may have relatively low bandwidth connectivity to the Internet, other departments and entities, e.g. some astronomy and high-energy physics communities, have very high bandwidth connectivity and we should target this diversity of groups and communities outside of Computer Science to add on as PlanetLab sites. We should also aggregisvely consider adding further geographic diversity by recruiting more sites in Latin America, Asia, Australia, Russia and so on. The connectivity characteristics of such diverse locations would certainly enhance the diversity on the PlanetLab. One possible approach to do this might be to examine various routing and topological data that are available from sites like Caida (http://www.caida.org), e.g. Skitter data, and identify specific sites within GREN that will expand PlanetLab diversity.

Focusing on the GREN for diversity has additional advantages. In general, the GREN is more *transparent* — there is more publicly available information about the connectivity of the GREN and more willingness to share information with the research community. Hence if we can find suitable diversity just within GREN, such an enhancement is certainly worth exploring further as we attempt to establish greater diversity within PlanetLab.

4.1 Acknowledgements

We would like to thank the following people whose comments on this work were extremely helpful — Randy Bush, Jon Crowcroft, Christophe Diot, Steven Hand, Gianluca Iannaccone, Z. Morley Mao, Andrew Moore, David Moore, Jennifer Rexford, Timothy Roscoe, Larry Peterson, James Scott, Colleen Shannon, and Richard Sharp. Thanks to Jose Augusto Suruagy Monteiro of UNIFACS, Sidney Lucena and Ari Frazao Jr. of RNP Brazil for helping us better understand the connectivity of the Brazilian research networks. We would like to thank Jeremey Stribling for collecting the PlanetLab round trip data and making it publicly available. In addition, this work would not be possible without the existence of public archives of interdomain routing data including Route Views and RIPE.

References

1. PlanetLab home page. http://www.planetlab.org.
2. Scriptroute Home Page. http://www.scriptroute.org.
3. University of Oregon Route Views Archive Project. http://www.routeviews.org.
4. M. Costa, M. Castro, A. Rowstron, and P. Key. PIC: Practical Internet Coordinates for Distance Estimation. In *24th IEEE International Conference on Distributed Computing Systems (ICDCS' 04)*, Tokyo, Japan, March 2004.
5. R. Cox, F. Dabek, F. Kaashoek, J. Li, and R. Morris. Practical, Distributed Network Coordinates. In *ACM Workshop on Hot Topics in Networks (HotNets-II)*, November 2003.
6. G. Hjaltason and H. Samet. Properties of Embedding Methods for Similarity Searching in Metric Spaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(5), May 2003.
7. H. Lim, J. Hou, and C. Choi. Constructing Internet Coordinate System Based on Delay Measurement. In *ACM SIGCOMM Internet Measurement Conference (IMC'03)*, Miami (FL), USA, October 2003.
8. Z. Morley Mao, David Johnson, Jennefer Rexford, Jia Wang, and Rany Katz. Scalable and Accurate Identification of AS-Level Forwarding Paths. In *INFOCOM '04*, 2004.
9. Z. Morley Mao, Jennefer Rexford, Jia Wang, and Rany Katz. Towards an Accurate AS-level Traceroute Tool. In *ACM SIGCOMM Technical Conference '03*, August 2003.
10. Ripe NCC. Routing Information Service Raw Data. http://abcoude.ripe.net/ris/rawdata.
11. T.S. Eugene Ng and H. Zhang. Predicting Internet Network Distance with Coordinates-Based Approaches. In *IEEE INFOCOM' 02*, New York, USA, June 2002.
12. L. Peterson, T. Anderson, and D. Culler. A Blueprint for Introducing Disruptive Technology into the Internet. In *ACM Workshop on Hot Topics in Networks (HotNets-I)*, October 2002.
13. M. Pias, J. Crowcroft, S. Wilbur, T. Harris, and S. Bhatti. Lighthouses for Scalable Distributed Location. In *2nd International Workshop on Peer-to-Peer Systems*, February 2003.
14. D. Spence. An implementation of a Coordinate based Location System. Technical Report UCAM-CL-TR-576, University of Cambridge, November 2003. Available at http://www.cl.cam.ac.uk/TechReports/UCAM-CL-TR-576.pdf.
15. C. Stover and M. Stanton. Integrating Latin American and European Research and Education Networks through the ALICE project. Document available at http://www.dante.net/upload/doc/AlicePaper.doc.
16. M. Szymaniak, G. Pierre, and M. van Steen. Scalable Cooperative Latency Estimation. Submitted for publication. Draft available at http://www.globule.org/publi/SCOLE_draft.html, December 2003.
17. L. Tang and M. Crovella. Virtual Landmarks for the Internet. In *ACM SIGCOMM Internet Measurement Conference (IMC'03)*, Miami (FL), USA, October 2003.