Empirical Study on the Evolution of PlanetLab

Li Tang[1,2], Yin Chen[1], Fei Li[1], Hui Zhang[1,2], Jun Li[2,3]
[1] Department of Automation, Tsinghua University, Beijing, China
[2] Research Institute of Information Technology, Tsinghua University, Beijing, China
[3] Tsinghua National Lab for Information Science and Technology, Beijing, China
{tangli03@mails., chenyin03@mails., li-f03@mails., zhanghui04@mails., junl@}tsinghua.edu.cn

***Abstract*—PlanetLab is a globally distributed overlay platform that has been increasingly used by researchers to deploy and assess planetary-scale network services. This paper analyzes some particular advantages of PlanetLab, and then investigates its evolution process, geographical node-distribution, and network topological features. The revealed results are helpful for researchers to 1) understand the history of PlanetLab and some of its important properties quantitatively; 2) realize the dynamic of PlanetLab environment and design professional experiments; 3) select stable nodes that possess a high probability to run continuously for a long time; and 4) objectively and in depth evaluate the experimental results.**

Keywords-PlanetLab; overlay; topology

I. INTRODUCTION

Due to their most complexity and heterogeneity, it is difficult to utilize simulation approaches for modeling and understanding the features of some worldwide distributed systems on the Internet [7]. While it is non-trivial to globally scatter computers to do actual experiments, PlanetLab [4] offers an open and common platform to conveniently achieve the purpose meanwhile without taking risk of harming the normal traffics.

The primary goal of PlanetLab is to support the design, evaluation and research of innovatively large-scale distributed techniques including algorithms, protocols, services and systems. Various advantages make PlanetLab distinctive and attractive. First, PlanetLab composes a planetary-scale network by means of collectively and collaboratively supported infrastructures. So far there are over 700 PlanetLab nodes distributed across more than 300 participating sites, and it still keeps expanding [4]. While its ultimate goal is to grow to 1000 geographically distributed nodes, PlanetLab with its current scale is difficult enough for most participating consortiums to build up alone by themselves. Second, the cooperation between academic, industrial, and government institutions, and the combination of edge sites, co-location and routing centers, and home users connecting to the Internet through ADSL or modems make PlanetLab diverse and heterogeneous, which is a crucial nature for imitating an environment similar to the Internet. Last but not least, PlanetLab provides a way to seamlessly migrate a distributed application from early prototype, through multiple design iterations, to a popular system that continues to evolve, as claimed in [6].

However, despite the fact that PlanetLab has been increasingly used and accepted in the literature, it may be unconvincing to take it as representative of the global Internet before understanding it clearly, especially as PlanetLab has always kept developing since its birth. [5] had investigated the inter-domain connectivity of PlanetLab and pointed out that PlanetLab might not be suitable to be taken as a representative of the Internet at that time, as about 85% PlanetLab nodes were located in so called “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.” Although it is widely accepted that PlanetLab has grown a lot since then, there is still a lack of decent study on how it has evolved and what is its recent status.

The main goal of this paper is to investigate the evolution process of PlanetLab, from its adolescence consisting of around 100 nodes dated back to early 2003, to its boom times including over 600 nodes until late 2005. By analyzing the historical data sustainedly collected by two projects, which have run for years to measure or monitor all the online PlanetLab nodes, the paper studies how the node number on PlanetLab has increased, how the geographical distribution expanded, and how the network topology developed. The reveal of the evolution of such properties is expected to help the researchers who are interested in doing experiments on PlanetLab to deeply understand it, make wisdom decisions on the node selection, and generalize objective and insightful conclusions.

The rest of this paper is organized as follows. Section II introduces the data set. Section III and IV present PlanetLab’s evolution process respectively in terms of the number of different types of nodes and the round trip times (RTTs) between them. Section V investigates the topological feature of PlanetLab, and Section VI makes conclusion.

II. DATA SET

The data used to conduct the study of PlanetLab’s evolution are gained from two projects, all-pairs-pings (APP) [2] and CoMon [3]. The remainder of this section briefly gives the necessary introduction on the means of the data collection and utilization.

A. APP Project

APP project deploys a set of scripts on each PlanetLab production node to perform periodically pings between every

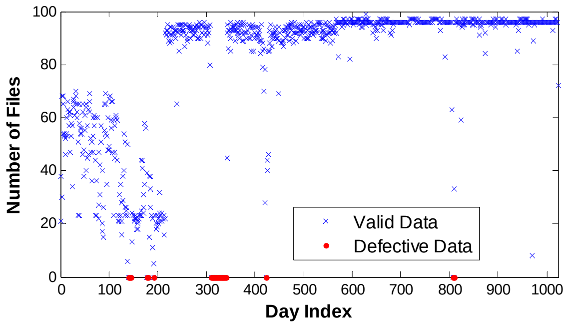

Figure 1. Daily Number of collected files in APP's data

node-pair. With an interval of 15 minutes, each node continuously pings each of all the other nodes until obtaining 10 RTT measurements or until timing out by 120 seconds per node-pair. A central server collects all the pings information once an hour, and for every interval the server archives the raw data collected, number of RTT measurements per node-pair, a list of PlanetLab node's IP addresses at that time, and a matrix of minimum/average/maximum RTT tuples of all the node-pairs. APP project checks the list of production node's IP addresses published by PlanetLab every hour to ensure its node list is up-to-date. As the number of nodes on PlanetLab keeps increasing, APP project reached the scaling limitations and was stopped on 12/1/2005.

This paper makes use of APP's data from 2/13/2003 to 12/1/2005, spanning 1023 days, to trace the growth of PlanetLab respectively in terms of the number of nodes and the statistics of RTTs. It is worth noticing that not all the data during such a long period are successfully collected, and it is possible there exist a few data with incorrect format and thus inadequate for analysis. Fortunately, after checking the data, we assure that the proportion of the missing or incorrectly formatted data is trivial and negligible.

APP project began to work stably with an archiving interval of 15 min since 9/15/2003, which leads the number of collected archives to be 96 or 97 each day. Fig.1. shows the number of archives collected every day during APP's lifetime, where the red points indicate there were totally 52 days without archives at all; but as can be seen, for most days seldom archives is missing.

B. *CoMon Project*

CoMon is a monitoring infrastructure providing a monitoring statistics for PlanetLab at both a node level and a slice level. Distinct from APP project, CoMon has archived the global data every 5 minutes since August 2004 and its node list is updated manually every day.

Despite CoMon record's various run time status of each PlanetLab node, this paper mainly concentrates on the names and IP addresses in order to understand the history of the mapping between each node's name and its IP address. In conjunction with the most recently detailed information of PlanetLab nodes given in [1], CoMon's data contribute to the investigation of how the PlanetLab nodes are distributed in the sites and geographically all over the world.

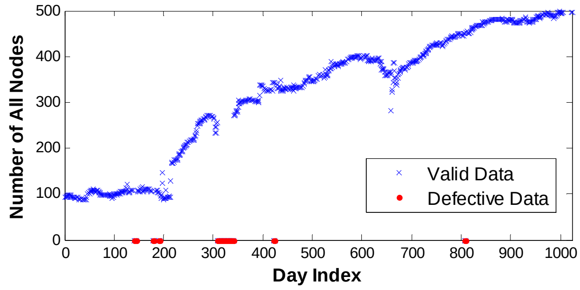

Figure 2. Daily number of all production nodes archived in APP's data

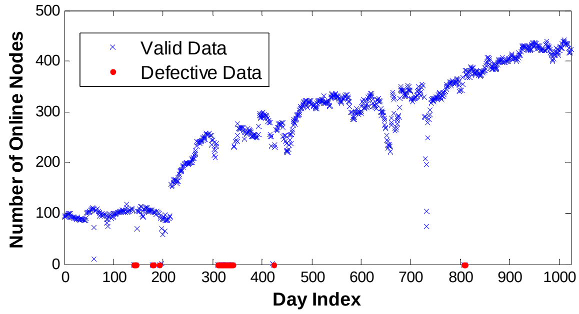

Figure 3. Daily number of online nodes in APP's data

III. SCALE EXPANSION

In this section, we reveal the expansion process of PlanetLab's scale by studying the evolution of the node numbers in APP's archives ordered chronologically.

A. *Number of Production Nodes*

PlanetLab has developed from 100 nodes to around 700 nodes in the last four years. Fig.2. reveals the actual expansion process of PlanetLab from 2/13/2003 to 12/1/2005. We can see a substantial growth of the number of nodes over time, which started to be noticeable around 8/31/2003. It is necessary to point out that the nodes in Fig.2. are limited to the production nodes, besides which, there are other kinds of nodes on PlanetLab, such as alpha nodes for system test or experiments requiring frequent restarts. Yet the fact that most nodes on PlanetLab are production ones makes APP's data still sufficient to reveal PlanetLab's expansion process.

B. *Number of Online Nodes*

Besides investigating the development of the number of all registered production nodes, researchers may show more interest to the online working nodes, as it is common for nodes to be occasionally unavailable on PlanetLab. Unfortunately, APP's data provide no explicit information about node status, which drives us to figure out an indirect way to achieve this purpose. For convenience of discussion, in the remainder of this paper we take all the RTT tuples in each APP archive as a delay matrix (DM). Given node *i* stands for the *i*th node in the archive's node list, we use *DM[i,j]* to denote the element of row *i* column *j*, which is actually the RTT tuple obtained when node *i* pings node *j*. Based on the DM, we define two types of ill nodes that are offline or unavailable due to network breakdown as follows:

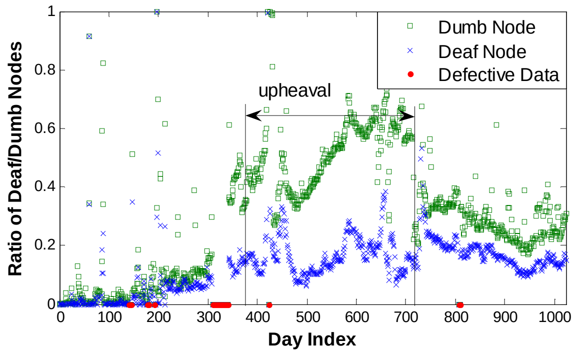

Figure 4. Daily ratio of the number of deaf/dumb nodes to that of all nodes

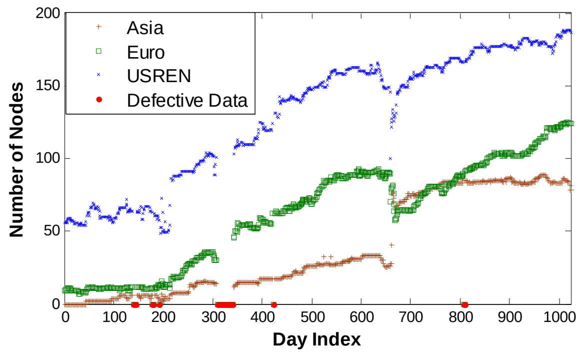

Fig.5. Geographical distribution of nodes every day in APP's data

- *Dumb Node* A node *i* is said to be dumb if and only if there is no data at all in the *i*th row of the DM, which means node *i* failed to submit data to the server in that interval.
- *Deaf Node* A node *i* is said to be deaf if and only if there is no valid data in the *i*th column of the DM, which means node *i* does not reply to all the probing messages.

If a node is deaf in an interval, it is determined to be offline at that time in the remainder of this paper. While it is theoretically possible that a node is online but deaf because the mediate networks between itself and every other node are breakdown, the probability is small enough to neglect. As can be seen, the development of the node number in Fig.2 and Fig.3 are almost the same, except the latter contains a bunch of descending segments. Moreover, it shows that there are more than 80% PlanetLab nodes online most of the time, which is optimistic to the perspective that around 30% PlanetLab nodes are offline at any time [9].

In some experiments or self-testing systems, a PlanetLab node is considered to be unavailable just because it fails to reply to a central server, which can surely lead to some misjudgement. Fig.4. depicting the ratio of deaf and dumb nodes to all nodes indicates the dumb nodes (nodes failing to reply to the server) are much more than the deaf ones (nodes unavailable to any other one), especially when the network status is unstable, e.g. the period between 2/27/2004 and 2/1/2005, as has been marked out in Fig.4. Actually, according to the ratio of the average number of dumb nodes and that of deaf nodes, the probability of misjudgement with the above method can be as high as 56%. Therefore, it is necessary to recommend the system designers to utilize more sophisticated method for the judgment of node unavailability.

C. *Nodes Locality*

To form a geographically distributed network platform, the nodes on PlanetLab are located all over the world. We determine the geographical location of the nodes according to their root domain names, excluding 20% of nodes that have a root DNS name of net, org, or com. Fig. 5. reveals the expansion of the nodes in three different continents where most PlanetLab nodes are located. In the beginning, most nodes were in the American Research and Education Network (USREN), and as time goes by an increasing proportion of nodes have been located in Asia and Europe.

Table.1. shows the characters of the deaf and dumb nodes respectively in the three continents. As can be seen, the reliability of nodes in USREN is significantly better than the other two. Therefore, the USREN nodes are highly recommended to the researchers who expect to do experiments in a domestic scale network.

TABLE I. STATUS OF THE NODES RESPECTIVELY IN THREE CONTINENTS

Rate	Asia	Europe	USREN
Deaf Node	0.169	0.133	0.107
Dumb Node	0.351	0.333	0.259

IV. RTT DISTRIBUTION

As the RTT between a pair of nodes is tightly related to their topological relationship and geographical distance, the distribution of RTTs thus is one of the important characteristics that reflect the evolution of PlanetLab.

A. *Daily Average RTT*

A daily average RTT (DAR) is defined to be the mean of all the average RTTs in the DMs of the APP archives within that day, which somewhat indicates the average network distances in terms of delay between every node-pairs on PlanetLab.

Fig.6. depicts the evolution progress of DARs along the 1023-day period, from 2/13/2003 to 12/1/2005. As can be seen, the DARs in the first 200 days (1st period marked in the figure) are nearly all smaller than 100 milliseconds, which implies PlanetLab used to be a domestic scale network at that time. The result agrees with Fig.5. that shows during that period around 90% nodes on PlanetLab are located within USREN. Afterwards, the DARs present an increasing trend on the whole for around 15 months (2nd period), ending with a drastic fluctuation. Still consistent with Fig.5., this is due to the gradually increased proportion of the European and Asian nodes on PlanetLab, and the fluctuation is mainly caused by the sudden participation of about 50 CERNET (China Education and Research Network) nodes placed in China at the year-end of 2004 and the accidental slump of the American and European nodes. Finally, the DARs become relatively steady

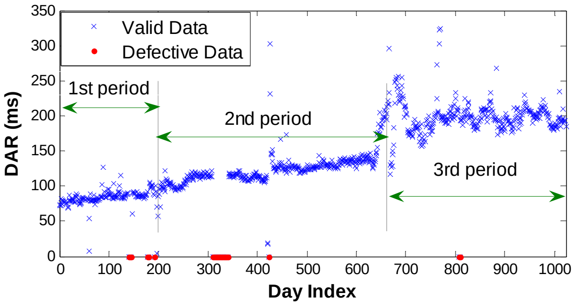

Figure 6. Daily average RTTs (DAR) from Feb. 2003 to Dec. 2005

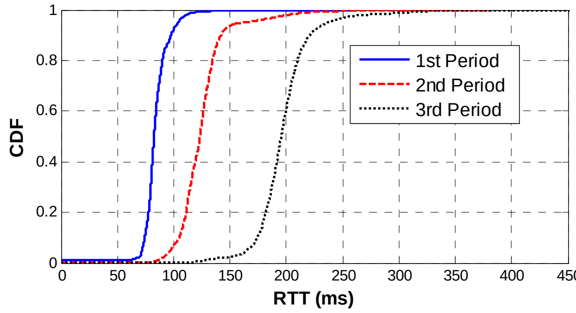

Figure 7. CDFs of RTTs in three different periods

around 200 milliseconds in the year 2005 (3rd period), which indicates that PlanetLab has grown to be a global scale network platform by that time and its DAR is likely to get relatively stable already.

B. RTT Distribution in Different Periods

To investigate the detailed characteristics of the RTTs in different periods, we plot the cumulative distribution functions (CDF) of the RTTs respectively in the three periods partitioned according to DARs in the last section, as shown in Fig.7. It clearly shows the differences among the RTT distribution characteristics in the three periods: in the 1st period, as most nodes are located in USREN, the RTTs among them are typically dozens of milliseconds; in the 2nd period, the participation of European nodes changes the RTT distribution, which implies that the RTTs between European and USREN nodes are generally larger than 100 and seldom exceeding 150 milliseconds; the RTTs in the last period show much larger diversity than the previous two, with the majority ranging from 120 to 300 milliseconds, which proves that since then PlanetLab has already been able to carry experiments requiring long-distance communication of world scale.

C. Variation of RTTs between Same Node-Pair

PlanetLab provides a scarce opportunity to study the long-term behavior of the RTTs between the same pair of nodes. To this end, we select two node-pairs that appear in APP's archives most frequently. The first pair of nodes are located in different continents, source in Inria Sophia Antipolis of France and destination in University of Michigan of USA; the second pair of nodes are both in the USA, source in Harvard University on Eastern Coast and destination in University of Utah near Western Coast. Fig.8. (a) (b) respectively show how the sampled RTTs with an interval of 15 minutes between each node-pair move with time. As can be seen, the RTTs between the same node-pair keep varying drastically even over short period all the time; a few extremely sharp peaks are even several times larger than the average; and there appears to be little, if any, alternation regularity of the RTTs. This significantly challenges the topological models that hypothesize static end-to-end delay between node-pairs. As such models have ever been widely used to design, optimize, and evaluate the algorithms and protocols of large-scale distributed systems that are expected to run on the Internet, it takes risk that their performance in real-life application may be quite different from the simulation results.

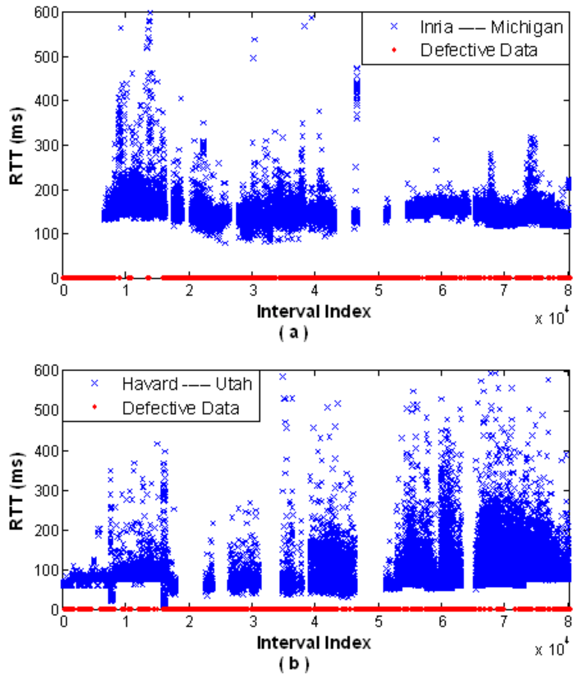

Figure 8. RTT variation between two specific node-pairs

V. TOPOLOGICAL FEATURE

The topological feature is studied with the *traceroute* data collected between every available node-pair on PlanetLab 7/23/2006. The *traceroute* command was configured by default, that is, using a 38-byte probing packet, 3 consecutive probes per hop, and restraining the hop count's maximum of 30.

A. Hop Counts of Paths

Totally 79969 routes (each record corresponds to a node-pair) were collected, out of which some failed to reach the destination node due to destination nodes offline or failures caused by packet loss, ICMP filtering, or the limit of maximum hop count. As 4930 routes (6.16%) exceed the *traceroute*'s default maximum of hop counts, it indicates that the operational diameter of PlanetLab has grown beyond 30 hops, which is comparable to or even larger than that of the Internet on 1995, according to the data Paxson collected [8]. This in turn suggests researchers doing experiments sending IP datagrams use large initial TTL values.

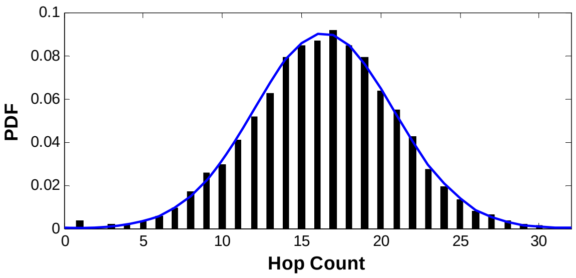

Figure 9. PDF of the hop counts of all routes

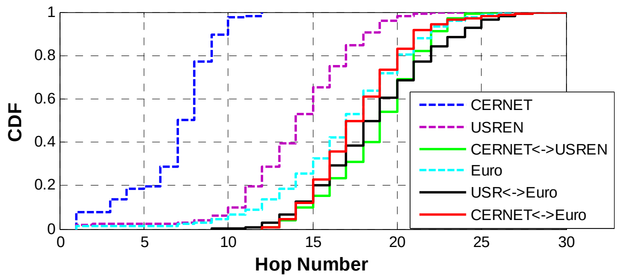

Figure 10. CDFs of the hop counts of intra/inter CERNET, USREN, and Europe routes.

The following results are based on the 73033 routes that successfully reach the destination nodes, involving 335 different nodes (every node ever acts as a destination, while only 240 nodes ever play a source). The nodes are scattered in 29 countries located in 5 continents, specifically, 113 nodes in North America, 105 in Europe, 29 in Asia, 6 in South America, 1 in Australasia, and 81 nodes with a root DNS of .com/.net/.org.

Fig. 9. shows the probability distribution function (PDF) plot of the hop counts of the investigated routes. As can be seen, 63% of the routes range from 13 to 20 hops, 85% from 10 to 22 hops, and only 1% are less than 5 hops or more than 28 hops. The average hop count is 16.4, which together with the Gaussian-like distribution of the hop counts approximates to the results given in [8], implying the similarity between PlanetLab's statistically topological property and that of the Internet.

The hop count of a route is highly related to the geographical relationship between the source and destination. As can seen in Fig. 10., most of the routes between a pair of nodes both located in CERNET have hop counts of no more than 10, while the routes of node-pairs within USREN mainly have 10 to 20 hop counts, a little smaller than those in Europe. This implies that the distances in terms of hop count between the PlanetLab nodes located in CERNET are generally shorter than that of the nodes located in USREN and Europe. The figure also indicates similar hop-count distribution of the routes between inter-continent node-pairs, all mainly ranging from 15 to 25.

B. *Different Significances of Routers*

The 73033 routes totally pass through 8772 different routers. Fig. 11. illustrates the distribution of the frequency of each router appearing in all the routes. The routers are arranged along the x-axis in the descending order of their frequencies that are represented by the y-axis. It shows that most routers are trivially placed on a few individual routes, for example approximately 76% appear in less than 100 routes and 37% less than 10 routes, while some routers are really much more critical, appearing in even over 8000 routes. The data also shows that the top 1% (87 routers) took up approximately 22% of the total transit time. This quickly escalates to the top 50% (4380 hosts) taking up 98% of the transit time.

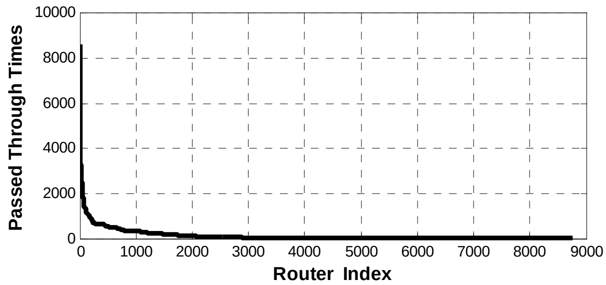

Figure 11. Times being passed through of each router

C. *Correlation Between RTT and Topological Heuristics*

This subsection aims to investigate the relationships between RTT and two topological heuristics, specifically the geographical distance and the length of common IP-address prefix. The correlation consists of two aspects. On the one hand, if two nodes are geographically distant to each other or have short common IP-address prefix, it usually implies they are also distant in the networks and thus suffer long RTT. For example, it is known that the speed of light in optic fiber is around 200 kilometers per millisecond and New York is more than 10,000 kilometers away from Beijing; then, only the propagation delay takes at least 100 milliseconds for a packet to make a round trip between two nodes respectively located in New York and Beijing. On the other hand, however, only these two heuristics are far from enough to accurately determine the RTT between two nodes.

The following results in this subsection are based on 29,373 node-pairs' average RTT, geographical distance and length of common IP-address prefix picked out from APP's data.

Fig. 12. illustrates the correlation of RTTs to geographical distances of node-pairs. We emphasize two important observations. First, the correlation is weak. Only given the geographical distances of several node-pairs, it is generally impossible to sort the node-pairs by their RTTs, not to speak of obtaining the RTT's exact value. Second, there is an upper bound of the speed of a packet traveling through PlanetLab. Given that the slope of the line in Fig. 12. is around 100, the fastest speed for a packet to travel through PlanetLab around the Earth surface is about 127 kilometers per millisecond, assuming the radius of the Earth to be 6371 kilometers.

Fig. 13. illustrates the statistics of the RTTs of the node-pairs having different lengths of common IP-address prefix. As can be seen, statistically, the RTTs of node-pairs decline with the increase of the length of common IP-address prefix. This is because of the techniques used in the IP address allocation aiming to reduce the routing table size of the core routers. The

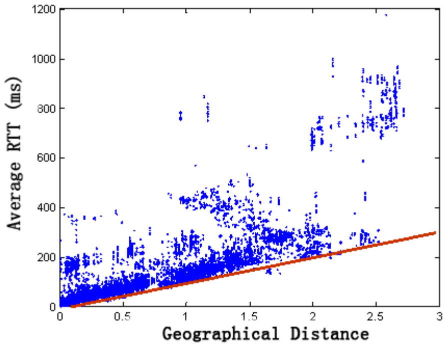

Fig. 12. Correlation between RTT and geographical distance that is measured by the spheral distance on the surface of the Earth that is considered as a sphere with radius of 1.

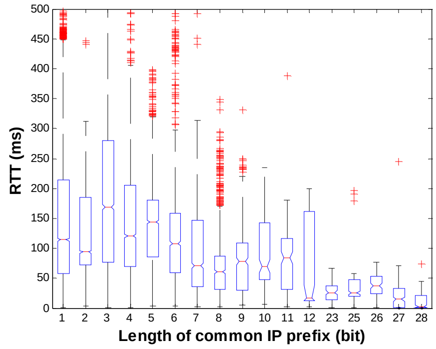

Fig. 13. Boxplot of the RTT to the length of common IP-address prefix, of which only those include more than 100 node pairs are analyzed in order to avoid misleading by individual cases.

Internet registries generally allocate IP addresses to Internet service providers (ISPs) by blocks represented by IP prefixes. For example, allocating the block of IP addresses 166.111.*.* to Tsinghua University means all nodes using IP addresses beginning with 166.111 are considered to be hosted in Tsinghua University. Fig. 13. also indicates that the RTTs of those node-pairs that have the same length of common IP-address prefix can vary drastically, especially when the prefix is shorter than 16 bits. This is due to the allocated IP address blocks are usually split into smaller fragments and used isomerically by ISPs. Nevertheless, the nodes with IP addresses having common prefix of more than 24 bits are usually in the same sub-network, and thus enjoy short RTT between each other.

VI. Conclusion

This paper quantitatively studies various characteristics of PlanetLab, revealing the evolution process since its adolescence consisting of around 100 nodes early 2003, until its boom times including over 600 nodes late 2005. The results show that: 1) while a majority of nodes are located in USREN in its early days, PlanetLab by now has already grown up to a world scale network platform in terms of geographical distribution, end-to-end delay and route hop count of the nodes; it has been capable of carrying experiments for Internet-oriented systems; 2) it is important to select proper nodes on PlanetLab, especially when the experiments require running continuously for a long time, as the node reliability is quite different from each other; generally, the USREN nodes have longer duration time in gear and are more preferable for domestic-scale experiments; 3) PlanetLab possesses a highly dynamic characteristics, which remind the researchers doing experiments on PlanetLab to carefully examine the status of PlanetLab nodes before jumping to some over optimistic conclusions; repeating experiments several times at different time is an effective way to discover unexpected results caused by the PlanetLab's anomaly.

Future work includes investigating the implication of PlanetLab's evolution process and topological features for the theory and design philosophy of highly available and scalable distributed systems.

Acknowledgment

This work is sponsored by NEC Laboratories China. The authors are grateful to the developers of APP and CoMon projects for sharing the data, and would like to thank all the members of FIT Security Lab for their discussion and advices.

References

[1] http://www.planet-lab.org/xml/sites.xml.

[2] "All-Pairs-Pings": http://pdos.csail.mit.edu/~strib/pl_app/.

[3] "CoMon": http://comon.cs.princeton.edu/.

[4] "PlanetLab": http://www.planet-lab.org.

[5] S. Banerjee, T. G. Griffin, and M. Pias, "The Interdomain Connectivity of PlanetLab Nodes," in Proceedings of the Passive & Active Measurement Workshop (PAM'04), April 2004.

[6] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, and M. Bowman, "PlanetLab: An Overlay Testbed for Broad-Coverage Services," ACM Computer Communications Review, vol. 33, 2003.

[7] S. Floyd and V. Paxson, "Difficulties in Simulating the Internet," IEEE/ACM Transactions on Networking, vol. 9, pp. 392~403, 2001.

[8] V. Paxson, "End-to-End Routing Behavior in the Internet," IEEE/ACM Transactions on Networking, vol. 5, no. 5, pp. 601~615, 1997

[9] L. Peterson, V. Pai, N. Spring, and A. Bavier, "Using PlanetLab for Network Research: Myths, Realities, and Best Practices" : http://www.planet-lab.org/PDN/PDN-05-028/, 2005.

[10] Y. Rekhter and T. Li, "An Architecture for IP Address Allocation with CIDR," RFC 1518, September 1993