On the Impact of Research Network Based Testbeds on Wide-area Experiments

Himabindu Pucha
School of ECE
Purdue University
West Lafayette, IN 47906
hpucha@purdue.edu

Y. Charlie Hu
School of ECE
Purdue University
West Lafayette, IN 47906
ychu@purdue.edu

Z. Morley Mao
Dept. of EECS
University of Michigan
Ann Arbor, MI 48109
zmao@eecs.umich.edu

ABSTRACT

An important stage of wide-area systems and networking research is to prototype a system to understand its performance when deployed in the real Internet. A key requirement of prototyping is that results obtained from the prototype experiments be representative of the behavior if the system were deployed over nodes connected to commercial ISPs. Recently, distributed testbeds such as PlanetLab and RON have become increasingly popular for performing wide-area experimentation. However, such testbeds typically consist of a significant fraction of nodes with connectivity to research and education networks which potentially hinder their usability in prototyping systems.

In this paper, we investigate the impact of testbeds with connectivity to research and education networks on the applications and network services so that such testbeds can be leveraged for evaluation and prototyping. Specifically, we investigate *when* the representativeness of wide-area experiments deployed on such testbeds is affected by studying the routing paths that applications use over such testbeds. We then investigate *how* the representativeness of wide-area experiments is affected by studying the performance properties of such paths. We further measure the impact of using such testbeds on application performance via application case studies. Finally, we propose a technique that uses the currently available testbeds but reduces their bias by exposing applications evaluated to network conditions more reflective of the conditions in the commercial Internet.

Categories and Subject Descriptors

C.2.1 [**Computer Communication Networks**]: Network Architecture and Design

General Terms

Measurement, Performance, Design, Experimentation

IMC'06, October 25–27, 2006, Rio de Janeiro, Brazil.

Keywords

Testbeds, Network characteristics

1. INTRODUCTION

The tremendous growth in the Internet has also seen a significant increase in research based on wide-area networks. A large body of work has concentrated on developing new distributed systems and network services for the wide area. For example, distributed systems such as DHTs (e.g., [20]), application-layer multicast (e.g., [9]), distributed storage (e.g., [21]), search (e.g., [29]), and file sharing (e.g., [35]) have been developed. Similarly, new network services to improve reliability (e.g., SOSR [7]), throughput (RON [1]), DNS performance (e.g., CoDNS [13]), web performance (e.g., CoDeeN [26]), and QoS (OverQoS [25]) have also been developed.

An important stage of all the above research is to prototype the distributed system and evaluate its behavior using relevant metrics in an Internet testbed. The specific metrics that are appropriate depend on the system under evaluation. A key requirement of prototyping is that results obtained from the prototype experiments should be representative[1] of the behavior if the system were deployed over nodes connected to commercial ISPs. This is desirable because: (1) More than 93% of all the ASes in the Internet belong to the commercial Internet and thus applications and network services should be evaluated under conditions prevalent in the commercial Internet. (2) The commercial Internet represents a much wider user-base and thus most distributed applications and network services are envisioned to be deployed primarily in the commercial Internet. Note that while there are applications intended for deployment in research networks (e.g., high performance grid computing), we are considering here research on general network services and applications intended to be deployed in the commercial Internet.

Ideally, more representative results can be achieved by evaluating the system using a wide-area distributed testbed whose nodes reside in the commercial Internet. The reality, on the other hand, is that obtaining any testbed of nodes can be a significant hurdle. Initially, researchers contacted friends and colleagues in other universities for access to machines to evaluate their systems (e.g., ESM [9]). Fortunately, the last few years have seen the development of shared wide-area testbeds such as RON [1, 19] and PlanetLab [15, 16] that facilitate wide-area experimentation. These testbeds

[1]The term representative is used loosely to denote similarity.

provide a distributed platform with nodes at several geographically distributed locations. Researchers use these testbed nodes as vantage points for measurements, as nodes that host their distributed system, or as end-hosts of an overlay network.

Although such testbeds are very popular and highly useful in the initial deployment and debugging of a system, the prototype performance/measurement results obtained from such a testbed will be more representative if the testbeds have nodes with commercial network connectivity. However, the common case for such testbeds is that a significant percentage of nodes have connectivity to the research and education networks. For example, when testbeds are formed by contacting colleagues [9], these colleagues are fellow researchers, typically working in a university environment. Such nodes belonging to educational institutions are usually connected to research and education networks. Similarly, 85% of nodes (typically hosted at educational institutions) in the popular PlanetLab testbed and 50% of nodes in the RON testbed have connectivity to the research and education networks. Thus, the connectivity of the currently available testbeds is not entirely commercial in nature, i.e., they are "mixed" testbeds containing a mixture of nodes with commercial and non-commercial connectivity. This hampers the representativeness of results obtained on the testbed if such testbeds are used for prototyping since non-commercial ASes (around 1400 in number) are a small fraction (less than 7%) of all ASes on the Internet.

Thus, in order to leverage these existing and available testbeds for prototyping distributed systems, it is essential to study the impact of such mixed testbeds on the representativeness of application performance. Such a study is significant since it enables system researchers to identify the limitations of these testbeds and hence leverage them suitably for their research. This paper, thus, focuses on the impact of mixed testbeds on the applications and network services being evaluated and prototyped over that testbed. Specifically, we investigate *when* the representativeness of wide-area experiments deployed on a mixed testbed is affected by studying the routing paths that applications use over such testbeds. We then investigate *how* the representativeness of wide-area experiments is affected by studying the performance properties of the non-representative paths. We further measure the impact of using such mixed testbeds on metrics used in distributed systems. Finally, we propose a simple technique that uses the currently available mixed testbeds but reduces their bias so that applications being evaluated are exposed to network conditions more reflective of the conditions when the applications are deployed in the commercial Internet. Another contribution of our work is a methodology to compare and calibrate various network performance metrics using mixed testbeds compared with commercial testbeds which can also be used to validate future testbeds for evaluating applications.

Our key findings are that:

- Current wide-area testbeds have a large fraction of nodes *worldwide* that use paths through research and education networks to communicate with each other, completely bypassing the commercial Internet.
- We find that the performance properties of these research and education networks are significantly different from the commercial Internet and this hampers the representativeness of network conditions experienced by applications evaluated on the testbeds. This impact is quantified through application case studies.
- Encouragingly, we find that the nodes with research and education connectivity can still be leveraged for experimentation since paths from these nodes to the nodes with commercial connectivity traverse a large fraction in commercial networks and were found to have similar distributions of performance properties.
- A simple overlay routing based solution can be useful to expose applications to network conditions more representative of the commercial Internet.

We note that this paper is not intended as a criticism of current testbeds such as PlanetLab and RON. In fact, we believe that these testbeds provide a valuable and highly useful service to the research community. Instead, our intent is to understand how best to leverage the existing testbed resources to perform representative wide-area experiments by assessing what aspects of an application's performance are affected by the testbed connectivity. Such an assessment has important implications since the testbed connectivity can affect the perceived performance of a wide range of distributed systems and networking experiments. Finally, while there is expected to be significant diversity due to different network tiers and a wide range of performance properties *within* commercial networks, this paper's focus is to study the coarse-grained distinction between commercial and GREN connectivity and not to explore the diversity within the commercial networks.

The rest of the paper is organized as follows. Section 2 discusses related work and Section 3 frames the problem by identifying the paths that application traffic can take in the current mixed testbeds. Section 4 studies the impact of using mixed testbeds on the topological properties of the routing paths taken by application traffic. Section 5 studies the impact of using a mixed testbed on the performance properties of the routing paths, and Section 6 further measures the impact on the performance of a set of applications. Finally, Section 7 advocates a technique for improving the representativeness of evaluation results using the current mixed testbeds and Section 8 concludes the paper.

2. RELATED WORK

Since deploying and evaluating new systems in operational networks is difficult, researchers in networks and distributed systems typically use discrete-event simulation, emulation, or live network experimentation on testbeds. While simulation tools provide control and repeatability, they sacrifice realism by abstracting many real world artifacts. Thus, many researchers currently evaluate their systems and services using emulation (real system on a synthetic network, e.g., EmuLab [27]) or wide-area testbed evaluation (real systems over real networks, e.g., PlanetLab [16], RON [19] testbeds). Newer proposals such as VINI [4] go even further than *overlay* testbeds such as PlanetLab by allowing routing protocols themselves to be modified and importing routing events from the real Internet into experiments. While emulation provides control and repeatability with more realism than simulators, wide-area testbeds can achieve better realism specifically with regard to live network conditions. This paper focuses on improving the realism of current wide-area

overlay testbeds that are widely used by systems researchers to evaluate new distributed systems and network services. Our contribution is in assessing and improving the realism of network conditions exposed by current wide-area testbeds.

The work in [3] first studied the interdomain connectivity of the PlanetLab testbed and argued that measurement research carried out using PlanetLab cannot automatically be taken as representative of the global Internet, since they reflect GREN characteristics rather than those of the global Internet; the term GREN is coined to refer to all the research and education networks collectively as the Global Research and Education Network. While this previous work pointed out an important fact by studying GREN-GREN connectivity, our paper takes this topic further and studies exactly how a "mixed" testbed impacts the distributed systems and network services evaluated over it across all traffic flow scenarios. Through measurement of a wide variety of performance properties, we quantify how GREN is different from the commercial Internet and whether it impacts applications. Finally, apart from a cautionary perspective on the use of such mixed testbeds, we advocate a technique to maximally leverage the testbed to provide results more representative of the commercial Internet where many distributed systems and network services are envisioned to be deployed.

It is well known [14] that Internet measurements heavily depend on measurement vantage points. Related to network measurements, wide-area application performance is also strongly influenced by the topological properties of the deployment locations. In our work, we focus on this latter problem, acknowledging the impact of network location on common network performance metrics such as delay and loss behavior. However, we attempt to quantify such impact from application's perspective.

3. APPLICATION TRAFFIC PATHS IN MIXED TESTBEDS

A key requirement in prototyping a distributed system using a testbed as an experimentation platform is to obtain performance/measurement results that are representative of the system when deployed on nodes that are part of the commercial service provider, i.e., the testbed nodes have commercial connectivity. When a testbed is composed of nodes that are connected to commercial ISPs only, the network conditions applications experience is more representative of the commercial Internet. However, popular testbeds such as PlanetLab are *mixed*, i.e., they have a significant fraction of nodes that have connectivity to research and education networks. GREN is a global network that connects various academic and research organizations by interconnecting many regional and national research networks. These include high-speed research backbones such as Abilene, GEANT, CANet, and regional educational networks such as CENIC. A schematic of many major networks part of the GREN network is shown in Figure 1. Current testbeds typically have a significant fraction of nodes *worldwide* connected to these depicted research and education networks.

The currently available mixed testbeds such as PlanetLab have been widely used for conducting a wide variety of distributed systems and networking projects. When such testbeds are leveraged to prototype a distributed system, the distributed system is typically deployed on a set of nodes

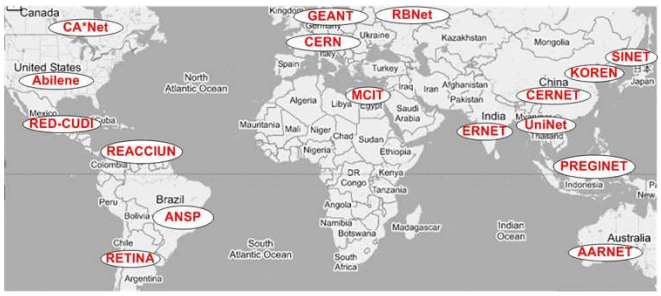


Figure 1: GREN Network schematic.

in the testbed. Typically, the application involves communication between pairs of testbed nodes. For example, an application-layer overlay multicast will involve communication between parent and child nodes in the multicast tree. The typical usage of the testbed can be classified according to the type of the networks the traffic flows in these distributed systems traverse. Consider a testbed with N nodes. These nodes can be divided into set C which consists of nodes with *only* commercial Internet connectivity and set G which consists of nodes with GREN connectivity. Note that nodes in set G may also have commercial Internet connectivity in addition to GREN connectivity. Consider a pair of nodes, A (sender) and B (receiver), that are part of the testbed and are communicating. The four traffic flow scenarios that can occur are:

- Case 1 - Both A and $B \in G$: This can occur for example in the CoDeeN [26] content distribution system on PlanetLab when a peer CDN node fetches content from another peer.
- Case 2 - $A \in G$ and $B \in C$: This can occur for example when an intermediate node from PlanetLab (belonging to G) is used to recover from path failure to a commercial Web server in the SOSR [7] system.
- Case 3 - $A \in C$ and $B \in G$: This can occur when a commercial Web server sends fresh content to a CDN node deployed in PlanetLab for example in the CoDeeN [26] system.
- Case 4 - Both A and $B \in C$: This can occur for example if both endpoints of a logical hop in a DHT happen to be nodes with commercial connectivity in PlanetLab [17].

Thus, distributed applications when deployed on current testbeds such as PlanetLab and RON can potentially use some combination of or all four traffic flow scenarios above during their operation. In the next sections, we study the impact of mixed testbeds on the representativeness of application performance by considering each of the above possible traffic flow scenarios.

4. IMPACT ON TOPOLOGICAL PROPERTIES OF NETWORK PATHS

In this section, we identify the traffic scenarios *when* the representativeness of applications is affected by the mixed testbed. In particular, we measure and analyze the routing paths used by the traffic scenarios to identify non-representative cases.

4.1 Methodology

To characterize the behavior in each of the four traffic flow scenarios mentioned in the previous section, we used the PlanetLab [15] testbed as an example of a mixed testbed consisting of nodes with non-commercial and commercial connectivity. The nodes in PlanetLab were divided into two groups: (1) nodes in G which have GREN connectivity and (2) nodes in C which have connectivity only through commercial ISPs. This breakdown is performed as follows: We obtained the origin ASN for the IP address of each PlanetLab node by identifying the ASN associated with the longest matching prefix to the IP. A list of prefix to ASN mappings was compiled by using the routing information from the BGP tables obtained from RIPE [18] and Route-Views [12]. We also compiled a list of GREN ASes by extracting all the routes announced by Abilene similar to in [3]. Further, to ensure correctness and completeness, we matched the above GREN AS list with another GREN AS list compiled by using the source field from a BGP RIB file in MRT format (obtained from Route-Views) and selecting out the data with source as Abilene, Indiana (ASN 11537). All nodes whose origin ASNs are contained in the GREN list are put in G and the remaining nodes are put in C.

We perform this study in two stages. In the first stage, only nodes within North America were considered. Here, after eliminating unresponsive nodes, the group G had 76 nodes while the group C had 11 nodes. In the second stage, we considered nodes all over the world. In this scenario, the group G had 155 nodes while the group C had 25 nodes. Interestingly not all nodes belonging to commercial organizations (e.g., HP Labs node *pli2.pa-3.hpl.hp.com*) belong in C. On the other hand not all university nodes belong to G (e.g., IIT Bombay, India node *planetlab1.iitb.ac.in*) since some do not yet have GREN connectivity. We then studied the routes taken by the traffic in Cases 1, 2, 3 and 4 described in the previous section to characterize their behavior.

To collect the routes along which traffic flows, we use the NANOG traceroute tool to obtain both AS-level and hop-by-hop traceroutes between any two nodes. The AS-level traceroute involves performing traceroute to the destination and then looking up the prefix of each hop in an AS-level registry. Currently, we configured the tool to use the *ris.whois.ripe.net* registry. We then classify these ASes into GREN and non-GREN ASes using the list of GREN ASes obtained from the BGP dumps.

Once the routes are collected, we parse the results to measure the fraction of the route in GREN (f_g). We measure f_g in two flavors: (1) f_{gH} - This is the ratio of the number of hops with in GREN to the total number of hops traversed by the traffic. A hop is assumed to be within GREN if both end points of the hop belong to a GREN AS. (2) f_{gRTT} - This is the ratio of the RTT experienced by the traffic within GREN to the total end-to-end RTT experienced by the traffic. We also measure the absolute number of hops within GREN (AH) and the absolute value of RTT within GREN (ARTT).

4.2 Case 1: Source and Destination in GREN

In Case 1 (both A and $B \in G$), when A communicates with B, it could either use the research backbone alone or the commercial backbone alone (since A may have commercial connectivity also) to route to B as depicted in the Figure 2(a). In the former case, the traffic may potentially flow

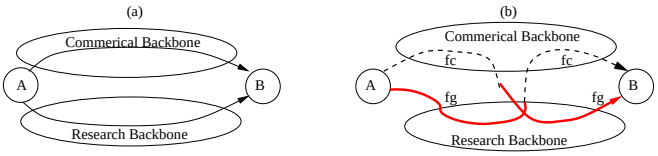


Figure 2: Illustration of source destination pairs.

entirely over GREN and thus have 100% of its route in a research network. In addition, A may also use a combination of research and commercial networks to route to B (as seen in Figure 2(b)). So when A communicates with B, it traverses a fraction (f_g) of its route in the research network and the remaining fraction f_c in the commercial network or vice-versa. Note that the values of f_g and f_c depend on the routing policy and are typically different.

For Case 1, we performed traceroutes among all source and destination pairs in set G for both North America and World nodes. For North America, this resulted in 5700 traceroutes out of which 531 traceroutes were unsuccessful while for World this resulted in 23,870 traceroutes out of which 4234 were unsuccessful. The traceroute failures occurred in this and other scenarios due to failure in receiving a reply from the intermediary routers. In this scenario, both the source A and the destination B have GREN ASes as their origin ASes. For each traceroute, we measure f_{gH}, f_{gRTT}, AH and ARTT.

The results of these measurements are shown in Figure 3. The results indicate that with few exceptions, A and B communicate with each other using routes that lie entirely within GREN, since for almost all traceroutes, 100% of the RTT lies within GREN. Importantly, this is true for both North American and World nodes indicating that research networks globally have such a routing policy. Interestingly, in a few exceptional cases, paths switch back and forth between research and commercial networks. For example, the paths *planetlab-1.eecs.cwru.edu* → *planetlab2.arizona-gigapop.net*, *planetlab-1.eecs.cwru.edu* → *planetlab1.iitr.ernet.in* and *planetlabone.ccs.neu.edu* → *planetlab2.arizona-gigapop.net* initially used a GREN network, switched to a commercial network (*gblx*, *Sprint* and *Level 3* respectively) and back again to a GREN network. We verified that this was not due to temporary failure of the GREN network.

In summary, with a few exceptions, *when source and destination have GREN connectivity, their communication entirely traverses research and education networks.*

4.3 Case 2: Source in GREN

In Case 2 ($A \in G$ and $B \in C$), the source A has GREN connectivity and the destination B has connectivity *only* via a commercial ISP. So when A communicates with B, it traverses a fraction (f_g) of its route in the research network before reaching B via the commercial network. However, since a GREN AS peers with other GREN ASes such as Abilene and regional networks (for example, California regional network (cenic.net), Illinois high school network (lincon.net)), the value of f_g is not apparent. The possible values for f_g are: (1) A can route within GREN until just before reaching B's commercial AS resulting in f_g close to 100%, (2) A exits into the commercial network right away from the source AS and uses its commercial connectivity to reach B, thereby causing $f_g = 0$, or (3) A routes for a fraction of its distance within GREN before exiting. Note that lower the value of f_g, the lesser the impact on the application performance and

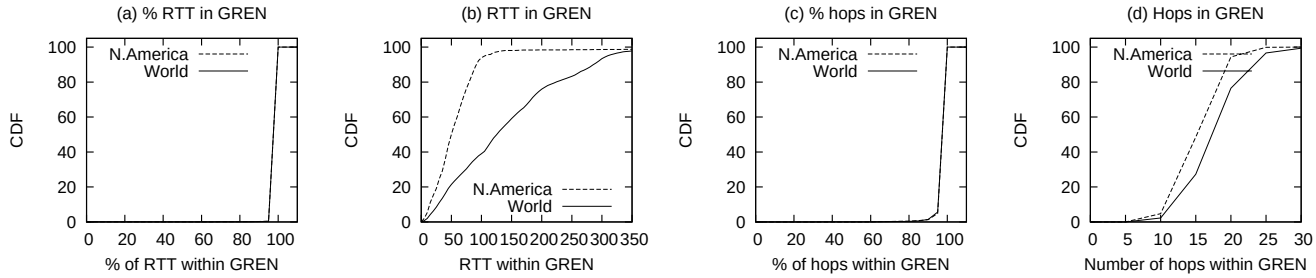


Figure 3: Study of routing paths in Case 1. The graphs depict the percentage and absolute RTT in GREN as well as the percentage and absolute network hops in GREN.

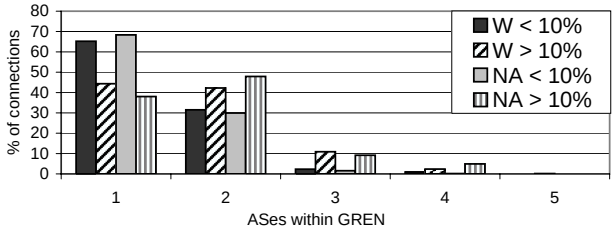


Figure 5: Case 2: AS hops inside GREN. "$W < 10\%$"/ "$NA < 10\%$" denotes all the paths in World / N. America with $f_{gRTT} < 10\%$ and so on.

higher the representativeness. Once again, we measure and understand the value of f_{gH}, f_{gRTT}, AH and ARTT.

We performed traceroutes from all nodes in G as sources to all nodes in C as destinations for both North America and World nodes. For North America, this resulted in 836 traceroutes out of which 79 were unsuccessful while for World this resulted in 3,879 traceroutes out of which 730 were unsuccessful. Figures 4(a)(b) depict the percentage RTT within GREN (f_{gRTT}) and the ARTT within GREN, respectively. Figure 4(a) shows that for up to 70% of connections in North America and 60% of connections in World, less than 10% of the RTT lies within GREN. We further analyze the behavior of these connections to answer the question that whether the packets originated from the GREN sources enter the commercial ASes directly from their origin ASes. We found that 65% of the routes with $f_{gRTT} < 10\%$ in both North America and World indeed exit into commercial networks from their origin ASes (Figure 5). However, another 30% of such routes (with $f_{gRTT} < 10\%$) traverse two GREN ASes before exiting into commercial networks while the remaining 5% of such routes traverse 3 or 4 GREN ASes. For example, node *planetlab-2.it.uu.se* when routing to destination *planet2.att.nodes.planet-lab.org* traverses 3 ASes belonging to "Swedish university networks" but with $f_{gRTT} < 2\%$. Thus, while some GREN nodes do not route traffic to commercial endpoints (e.g., Abilene, CANet), since in practice, almost every one of the GREN nodes is also connected to regional and other research networks that have dedicated links and hence are part of the GREN, many GREN nodes use these GREN networks to route to the destination, thereby traversing more than one GREN AS.

The following reasons account for the connections with $f_{gRTT} > 10\%$: (1) The percentage RTT within GREN also depends on the absolute value of RTT to the destination. The closer the destination is to the source, the larger the percentage RTT within GREN. This contributes to some connections having a larger fraction of RTT within GREN. (2) Some GREN nodes route to specific commercial nodes through paths entirely in GREN. For example, GREN nodes located in California route to *planet3.berkeley.intel-research.net*, belonging to Intel Research, Berkeley, using the regional California network (cenic.net) and the entire path lies within GREN. Such connections contribute to the small fraction of connections with 100% of RTT within GREN. (3) Some routes traverse multiple GREN ASes before exiting to commercial networks and hence have larger values of f_{gRTT}. For example, when *planetlab1.flux.utah.edu* routes to *planet2.att.nodes.planet-lab.org*, the route traverses 3 unique GREN ASes before exiting to commercial network, resulting in $f_g = 37.5\%$. Figure 5 depicts the breakdown of how many ASes a packet travels before it exits the GREN network for $f_{gRTT} > 10\%$. Despite a major fraction of nodes exiting GREN directly from their origin AS, some nodes traverse other GREN ASes like regional networks before exiting GREN. (4) We also observe that a few connections travel long distances in a single GREN AS before exiting, resulting in larger f_{gRTT}.

Interestingly, very few paths (e.g., *planetlab2.eecs.iu-bremen.de* → *planetlab1.singapore.equinix.planet-lab.org*) also exhibit switching from GREN to commercial to GREN and back to commercial networks. Finally, Figure 4(b) depicts that up to about 100% of connections in N. America and 80% of connections in the World have an absolute RTT of less than 10ms within GREN. Thus, the absolute RTT within GREN is negligible for these connections.

Figures 4(c)(d) depict the number of router hops taken until the route exits the GREN network. Figure 4(c) shows that for up to 81.8% of connections in N. America and 74.7% of connections in World, up to 40% of hops lie within GREN. Once again, the closer the destination, the lower the number of total hops to it, and the higher the percentage of hops within GREN. Also, the percentage of hops in GREN is different from that of RTT since each hop contributes a different RTT. For instance, typically, at the source and the destination, a route travels several hops with an RTT of approximately 1 ms. It is the hops in the middle of the path that contribute the large fraction of RTT. The absolute number of hops in Figure 4(d) indicates that up to 50% of connections travel less than 7 hops before exiting GREN network while the number of hops traveled in GREN ranges uniformly from 1-10.

We conclude that for a significant fraction of connections, packets from a GREN to a commercial node stay within GREN for a small fraction of the time. This suggests the use of a predominantly "hot-potato" routing policy in GREN networks when routing to nodes in commercial networks.

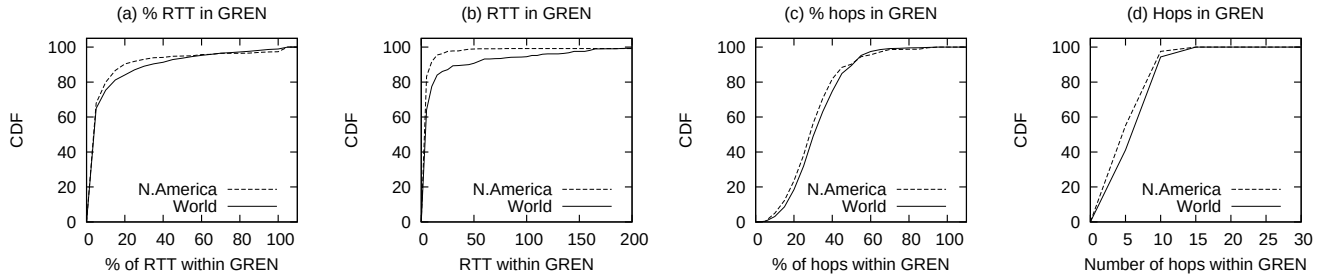


Figure 4: Study of routing paths in Case 2. The graphs depict the percentage and absolute RTT in GREN as well as the percentage and absolute network hops in GREN.

4.4 Case 3: Destination in GREN

We now examine the path of a commercial node reaching a GREN node. In Case 3 ($A \in C$ and $B \in G$), when A communicates with B, it needs to reach B by traversing a fraction (f_c) of its route in the commercial network. Alternatively, A may reach B using its commercial connectivity directly, thereby causing $f_c = 1$. The higher the value of f_c, the lesser the impact on the application performance. Note that $f_c = 1 - f_g$. Hence, to understand the routing from A to B, we again evaluate f_{gH}, f_{gRTT}, AH and ARTT.

We performed traceroutes from all nodes in C as sources to all nodes in G as destinations for both North America and World nodes. For North America, this resulted in 836 traceroutes out of which 163 traceroutes were unsuccessful while for World this resulted in 3,879 traceroutes out of which 724 were unsuccessful. Figure 6(a) depicts the percentage RTT within GREN (f_{gRTT}) while Figure 6(b) depicts the ARTT within GREN. Figure 6(a) shows that for up to 70% of connections in North America and 65% of connections in World, less than 10% of the RTT lies within GREN. Among these connections, 65% of them traverse a single GREN AS in both North America and World (Figure 7). The remaining connections traverse 2-3 ASes within GREN networks.

The following reasons explain why $f_{gRTT} > 10\%$ for the remaining connections : (1) Similar to in Case 2, connections with small end-to-end RTT contribute to larger f_{gRTT}. (2) Once again, certain commercial nodes use GREN ASes to route to certain GREN destinations. For example, for commercial node *planet3.berkeley.intel-research.net*, the entire route to some other GREN nodes lies within GREN. These connections have 100% of their RTT within GREN. The destinations in this case include GREN nodes in California which are routed to through GREN (*cenic*). This behavior is also observed in Case 2. However different from Case 2, apart from the California GREN nodes, this node connected to nodes across the pacific ocean (e.g., in Japan, Australia and Taiwan) using GREN (*pacificwave* non-profit network). This was also observed for other commercial nodes in California such as *planetlab-1.sjce.nodes.planet-lab.org*. Interestingly, the backward route from the GREN nodes in Japan, Australia and Taiwan to commercial nodes in California actually traversed a commercial network. Thus peculiar asymmetry can exist, i.e., the forward path uses GREN and the backward path uses commercial networks. (3) Similar to in Case 2, paths in Case 3 also traverse multiple ASes *within* GREN (Figure 7). The number of paths taking greater than 1 AS hop in GREN are also similar (31% in case 2 and 30% in Case 3). (4) Finally, certain paths traverse longer hops within GREN after they exit the commercial network even if the number of GREN ASes traversed are small. For example, when node *planet2.att.nodes.planet-lab.org* communicates with *planetlab1.iis.sinica.edu.tw*, 190ms of RTT is within GREN out of a total of 262ms.

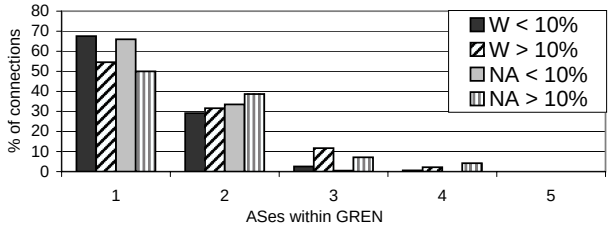


Figure 7: Case 3: AS hops inside GREN. "$W < 10\%$"/ "$NA < 10\%$" denotes all the paths in World / N. America with $f_{gRTT} < 10\%$ and so on.

Figure 6(b) depicts that close to 85% of connections in N. America and 60% of connections in the World have an absolute RTT of less than 10ms within GREN. These percentages are slightly lower than those in Case 2 suggesting that routes from nodes in G exit faster from GREN when going to a node in C while the routes from nodes in C enter earlier into GREN when going to a node in G. This again suggests asymmetry exists in the way GREN is utilized in forward and backward paths.

Figures 6(c)(d) also depict the number of router hops taken until the route enters the GREN network. Similar to Case 2, Figure 6(c) shows that for up to 78.6% of connections in N. America and 68.5% of connections in World, up to 40% of hops lie within GREN. Once again, the closer the destination, the lower the number of total hops to it, and the higher the percentage of hops within GREN. Also, the percentage of hops is different from the RTT since each hop contributes a different RTT. The absolute number of hops in Figure 6(d) indicates that up to 80% of connections travel less than 7 hops before entering the GREN network while the number of hops in the GREN network range from 1-15 (slightly higher than in Case 2).

We, thus, conclude that packets from a commercial node to a GREN node typically stay within GREN network for a small fraction of time. This suggests the use of a predominantly "late-exit" routing policy (common for customer networks) when commercial networks communicate to a GREN node in contrast to the "hot-potato" routing policy used when a GREN node sends packets through its commercial upstream provider networks. Further, the observations made in Case 3 are largely similar to that in Case 2. Another interesting observation from both Case 2 and Case 3 is that when a GREN node communicates with any commer-

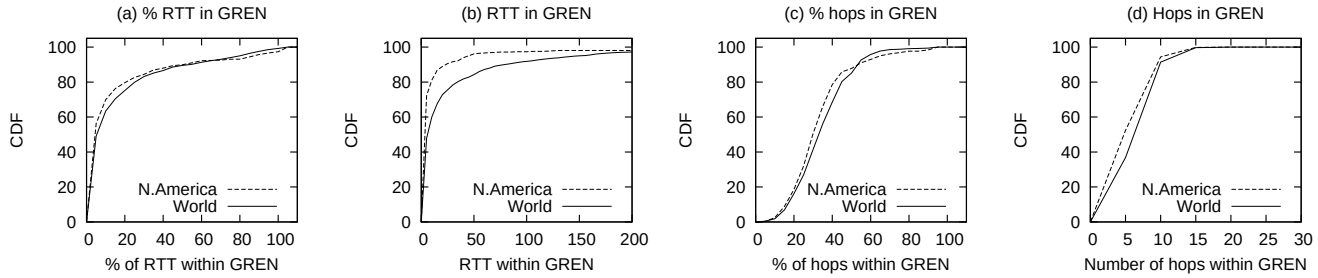


Figure 6: Study of routing paths in Case 3. The graphs depict the percentage and absolute RTT in GREN as well as the percentage and absolute network hops in GREN.

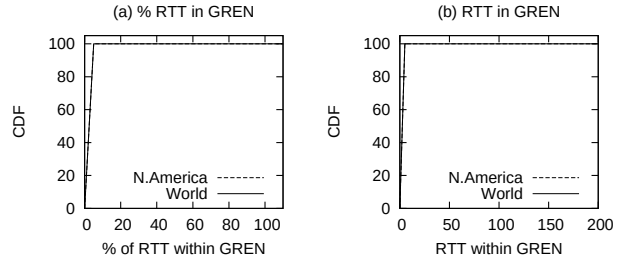


Figure 8: Study of routing paths in Case 4. The graphs depict the percentage and absolute RTT in GREN.

cial node, the traffic from the GREN node exits the GREN (i.e., the last hop within GREN) from one of a deterministic number of exit points. We refer to them as *markers*, which are effectively BGP peering locations. Thus, each source GREN node can be associated with a set of markers at which traffic enters commercial networks. Similarly, when any commercial node communicates with a GREN node, the traffic enters the GREN from one of a deterministic number of entry markers. Also, entry and exit markers can potentially overlap.

4.5 Case 4: Source and Destination in Commercial Networks

In Case 4 (both A and $B \in C$), the situation is similar to a testbed consisting of nodes only from the commercial Internet and thus will not impact the application performance. However, we wanted to study this case to make sure there are no corner cases and that the traffic does indeed flow over commercial networks for all flows. For Case 4, we performed traceroutes from all nodes in C as sources to all nodes in C as destinations for both North America and World nodes. For North America, this resulted in 110 traceroutes out of which 22 traceroutes were unsuccessful while for World this resulted in 600 traceroutes out of which 100 were unsuccessful.

The results depicted in Figure 8 confirm that almost all paths travel entirely outside GREN and in the commercial Internet (RTT in GREN is close to 0ms for almost 100% of connections). A rare exception is *planet3.berkeley.intel-research.net* which uses GREN when routing to commercial nodes *in California* such as *sanfrancisco.planetlab.pch.net*.

4.6 Summary

In summary, the measurements presented in this section indicate that when applications using a mixed testbed result in communication between two GREN nodes[2], the entire traffic flows over research and education networks, completely bypassing the commercial Internet. This can potentially hamper the representativeness of network conditions experienced by applications and potentially reduce the usefulness of utilizing GREN nodes for experiments. However, on a positive note, measurements also indicate that in Case 2 and Case 3; although the traffic in both cases traverses GREN, the absolute RTT and the fraction of RTT within GREN is negligible. Thus, when applications using the mixed testbed result in communication between a GREN node and a commercial node, the traffic predominantly flows over the commercial network. This fact potentially allows a careful use of GREN nodes in experiments.

5. IMPACT ON PERFORMANCE PROPERTIES OF NETWORK PATHS

The previous section has shown *when* the representativeness of application traffic flows is affected, in particular, the flows between GREN nodes bypass commercial networks. Since the performance of any distributed application is directly dependent on the network performance properties of paths over which its traffic flow, we further study the impact of the testbed on these performance properties to understand *how* it affects the representativeness.

We proceed in two steps. First, we study how different the performance properties of the GREN flows are compared to flows over commercial networks. Second, we study whether the performance properties of GREN-commercial and commercial-GREN paths also stay representative. Although such paths appear to be representative as they travel largely over commercial networks, the nature of the GREN-commercial network peering can potentially be different from commercial-commercial peering enough to affect the performance properties of such paths even if they travel largely in commercial networks.

5.1 Methodology

We study the following six performance properties of paths in this section:

Round Trip Time (RTT) We measured the round trip time between nodes representing Cases 1-4 using TCP ACK /RST based probes to high numbered ports since they better reflect the RTT experienced by TCP packets flowing through routers and are not filtered like ICMP probes. Each

[2]The probability of communication between a pair of GREN nodes is high due to the typically larger fraction of GREN nodes in mixed testbeds (e.g., 85% in PlanetLab).

measurement uses 25 probes and was performed by each node 3 times a day. We recorded the average, standard deviation and minimum RTTs (similarly as in [3]).

Path Loss We report loss rates measured by the `tulip` tool [10]. The overall conclusions were also confirmed through the `BADABING` tool [22]. The data was averaged over 10 measurement runs.

Throughput We measured the TCP throughput achieved between nodes representing Cases 1-4 using `Iperf` [31]. Iperf is a widely used tool used to measure achievable TCP throughput and measurements between PlanetLab nodes are available [32]. We found the average and standard deviation of throughputs for 90 traces collected over a 3 month period.

Available Bandwidth The average and standard deviation of the end-to-end available bandwidth is measured using `Spruce` [24] since it was shown to be more accurate than other tools designed for the same purpose in [24]. The data consisting of 10 different traces was collected from S^3 [33, 28], that provides a sensing service for large-scale distributed systems such as PlanetLab.

Capacity The average and standard deviation of the bottleneck capacity is measured using `pathrate` [6]. The data consisting of 10 different traces was collected from the S^3 [33, 28] service. Note that the capacities observed do not directly correspond to rate plans provided by ISPs due to measurement noise. However, the trends observed in the capacities measured are sufficient for our problem scope.

Bottleneck Location The bottleneck location is measured using the `pathneck` [8] tool. The measurements presented are based on a single run of the tool.

These properties are studied using a list of 22 nodes in PlanetLab, 11 each belonging to G and C. We choose the nodes in pairs (one each from G and C) such that each pair has a similar geographic location. The nodes chosen are both in the USA and outside USA. The locations of the nodes and the organization they belong to are listed in Table 1. The C2C (Case 4) connections are among the 11 commercial nodes while the G2G (Case 1) connections are among the 11 GREN nodes. Thus, in total, there are 110 G2G paths and 110 C2C paths for which we compare performance properties.

The nodes chosen do provide significant diversity in their outbound paths. For the 22 nodes used in the study, 15 unique commercial ISPs are used for Internet connectivity. Some larger ISPs such as Level3, AT&T were used by multiple nodes. However, the actual paths used in terms of physical hops are typically different for each node even if they use the same ISP. This is due to their geographical separation. In addition, even if nodes use the same ISP, the AS-level path traversed by them to a common destination may still be different based on routing preference.

5.2 The Difference between Flows in GREN and Commercial Networks

We first compare the performance properties of flows in Case 1 (G2G) and Case 4 (C2C). Despite the fact that the commercial and GREN networks are physically separate (and thus different), we still wish to measure whether there exists significant difference in the performance properties of paths in these networks, enough to impact the evaluation of applications over the testbed.

Location	GREN	Commercial
New Jersey 1	Princeton U.	ATT Labs
Oregon	U. Oregon	CTG ISP
California 1	Berkeley	IRL, Berkeley
New Jersey 2	Columbia U.	NEC Labs
California 2	Stanford U.	HP Labs
Pennsylvania	U. Pittsburgh	IRL, Pittsburgh
Iowa	Iowa State U.	ATCORP
Washington	U. Washington	IRL, Seattle
Cambridge, UK	U. Cambridge	IRL, Cambridge
Warsaw, Poland	Warsaw U. Tech	TP Group
Amsterdam	Vrije U.	PlanetLab-AMST

Table 1: Locations of GREN and commercial nodes.

5.2.1 RTT

Figure 9 shows the distribution of RTTs that will be experienced by applications over G2G and C2C paths. The first two figure depicts the average RTT experienced for a path along with the standard deviation of RTT observations recorded for that path. The paths are sorted according to their RTT. The results show that G2G and C2C RTTs have different characteristics despite similar geographic locations of the nodes considered. While the maximum average RTT in G2G can be up to 200ms, the maximum average RTT can be up to 300ms in C2C. The figures also show that average RTTs of individual paths in G2G are slightly lower for most samples than in C2C. The average RTT of a C2C path was 24.4% higher than that of a G2G path. Also, while 37% of the RTTs in G2G paths were more than 100ms, 50% of the RTTs on C2C paths were more than 100ms. C2C RTTs are also more unstable. While only 2 paths in G2G exhibit high variance, approximately 21 paths in C2C had high variance. The last figure also depicts the minimum RTT recorded for G2G and C2C paths. As expected these distributions are similar since they tend to reflect the real distance between nodes and we chose these nodes in similar locations. However, the dynamic variations of RTT is larger on C2C paths than on G2G paths.

In conclusion, the RTTs experienced under Case 4 can be different (typically higher average and variance) from that experienced under Case 1. These increases are likely due to path inflation from routing policies [23] of commercial ISPs.

5.2.2 Path Loss

Path loss is an important property that affects the throughput achieved by TCP-friendly application flows that are commonly used by several applications ranging from application layer multicast, DHTs and data dissemination systems.

The `tulip` tool [10] was used to measure loss rates for the G2G and C2C paths and the resulting loss rate distributions are shown in Figure 10. These results show that while the loss rate is close to 0 for 58% of G2G paths, it is close to 0 for 45% of C2C paths. The average loss rate for G2G paths was 0.059 while the loss rate for C2C paths was 0.078. In summary, we believe that overall G2G paths exhibit lower losses than C2C paths. Commercial networks carry more traffic and this could be one of the reasons for this observation.

5.2.3 Available Bandwidth and Capacity

Many high throughput demanding applications such as those in data dissemination can vary in performance depending on the available bandwidth and capacity available

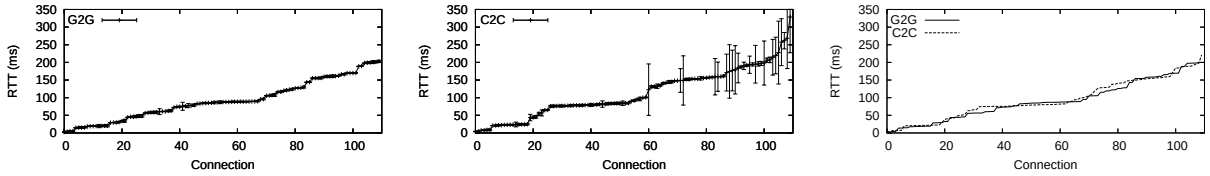


Figure 9: RTT measured for G2G paths and C2C paths using TCP ACK/RST packets. The last figure depicts the minimum RTT recorded.

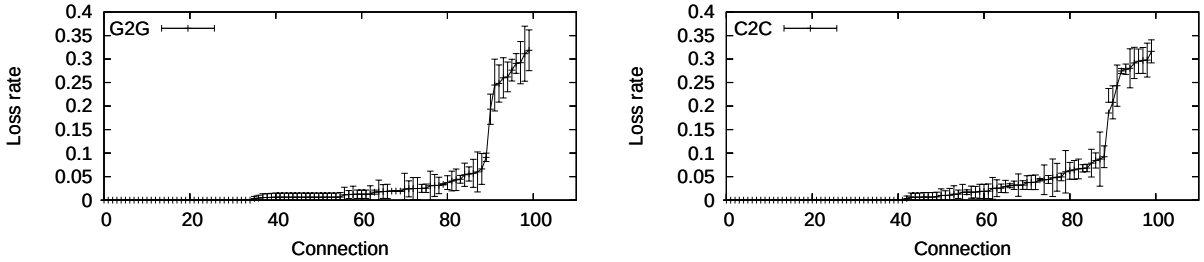


Figure 10: Path loss rates measured for G2G and C2C paths.

on Internet paths. Figure 11 shows the average and standard deviation of the available bandwidth between the pairs of nodes in Case 1 and Case 4 found using the `Spruce` tool. The results show that the average available bandwidth on G2G paths are significantly higher than those on C2C paths. For example, the maximum average available bandwidth measured on a C2C path was 50,000 Kbps while more than 80% of the G2G paths had more than 50,000 Kbps available bandwidth on average. Overall, the average available bandwidth was 600% higher on G2G paths than that on C2C paths. Additionally, while 95% of G2G paths had an available bandwidth greater than 10Mbps, only 37% of C2C paths had available bandwidth more than 10Mbps.

Similar trends were seen when measuring the capacity of G2G versus C2C paths. Figure 11 also shows the capacity between the pairs of nodes in Case 1 and Case 4 measured using the `pathrate` tool. The results show that C2C paths have significantly lower capacity than G2G paths. Overall, the average capacity on G2G paths was 532% higher than that on C2C paths. Additionally, while 95% of G2G paths had capacity greater than 50Mbps, only 25% of C2C paths had available bandwidth more than 10Mbps.

5.2.4 Bottleneck Location

To investigate whether Case 4 (C2C) paths having lower capacity and available bandwidth than Case 1 (G2G) paths is an artifact of the commercial nodes having lower access bandwidth or a general observation about the commercial network paths, we used the `pathneck` tool to locate bottlenecks on all 110 C2C paths.

Table 2 shows the top 3 choke points and their confidence levels as reported by pathneck as well as whether the choke point resides in the network (I) or at the access link A. The results show that for 64.8% of the paths considered, the main choke-point (CP1) lies in the network rather than the access link and overall 66.4% of the paths have at least one of the top three chokepoints in the network. Thus the previous results about capacity and available bandwidth are primarily because Case 4 paths have lower capacity and available bandwidth than Case 1 paths and that a majority of the bottlenecks are in the core of the network.

5.2.5 Throughput

Finally, we measure the actual TCP throughput achieved over all 110 G2G and all 110 C2C paths to compare the final impact all network properties can collectively have on application throughput. Figure 12 shows the TCP throughput between the pairs of nodes in Case 1 and Case 4 measured using `Iperf`. Again we find that C2C paths on average have lower throughput than G2G paths. Note that while the G2G paths have large available bandwidth and capacity, the obtained TCP throughputs are lower in scale since the actual connections may have large RTTs, path losses may occur or the PlanetLab node may be overloaded; all of which can reduce the final TCP throughput achieved. However, the important fact is that the TCP throughputs achieved are higher than on C2C paths. Overall, the average throughput obtained on G2G paths was 127% higher than that on C2C paths. Additionally, while 70% of G2G paths had a throughput greater than 1500 Kbps, only 30% of C2C paths had a throughput greater than 1500 Kbps.

5.3 Representativeness of Case 2 and Case 3

The previous section showed that Case 1 and Case 4 paths are very different from each other in performance properties, and hence the use of G2G paths for evaluating application performance can make results less representative. It remains to be seen whether we can still leverage the nodes belonging to G to perform experiments and this depends on how representative Case 2 and Case 3 paths are. Thus, in this section, we measure the representativeness of Case 2 and Case 3 traffic.

The results are shown in Figure 13. The figure compares performance properties for all 110 C2C, G2G, C2G and G2C paths. Figure 13(a) shows the RTT for all four types of paths. The curves show that C2G and G2C RTTs are similar in nature to each other and to C2C except in a few cases and are slightly higher than G2G RTTs for a majority of paths. More importantly, Figures 13(d),13(e) show that the available bandwidth and capacity on C2G, G2C and C2C paths are similar to each other and all lower than those of G2G paths. One anomaly seen is that in Figure 13(d) some of the G2C paths have higher available bandwidth than the

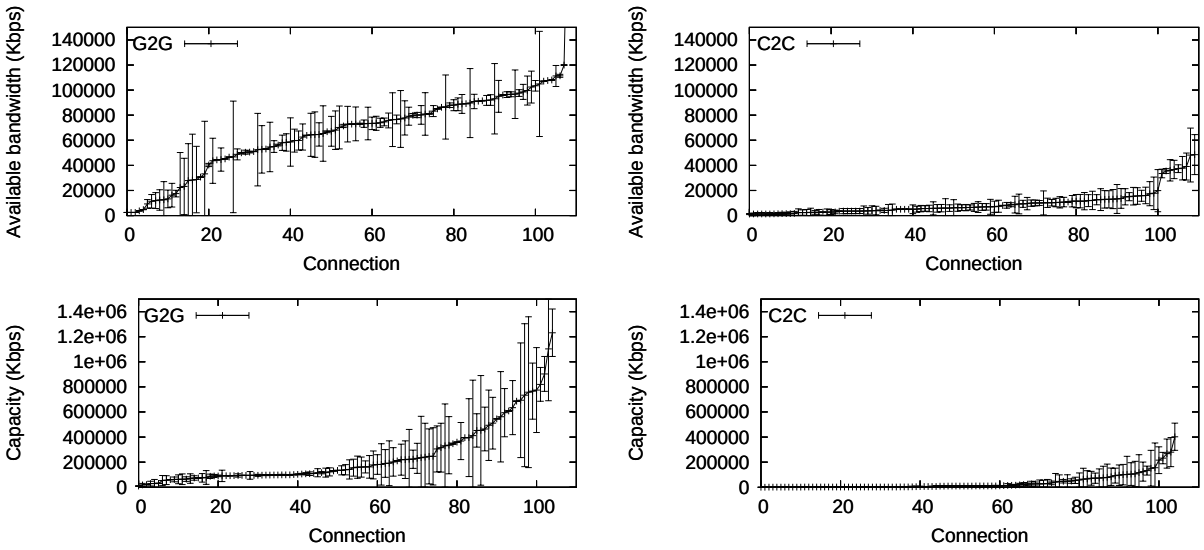


Figure 11: Available bandwidth and capacity measurement for G2G and C2C paths.

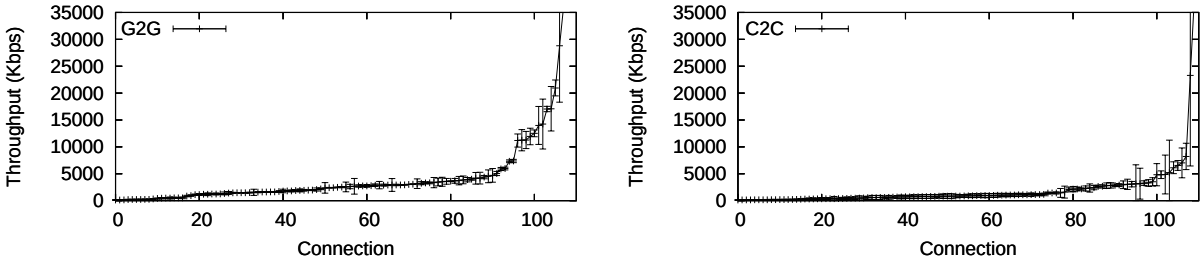


Figure 12: TCP throughput measurements for G2G and C2C paths using `Iperf`.

C2G and C2C paths. We determined that this was due to the access link of one of the nodes (ATT) being limited to approx. 10Mbps. While this limited the bandwidth of C2C and C2G paths, the corresponding paired GREN node at Princeton did not have this limitation and thus the G2C paths have higher bandwidth. Figure 13(c) also shows that the loss rates of G2C and C2G paths are higher than those of G2G paths.

Finally, Figure 13(f) shows the TCP throughput achieved on each of these paths which most closely indicates how distributed systems and network services will perform over the testbed and takes into account collectively all path properties like RTT, loss and available bandwidth. The results show that the throughput achieved on G2C, C2G and C2C paths for most of the measured paths are similar, while G2G paths have higher throughputs.

5.3.1 Relative Change Metric

We also measure the relative change of G2C and C2G paths compared to the corresponding C2C and G2G paths. The relative change for a G2C path compared to a G2G path for metric I is defined as $\frac{|I_{G2C}-I_{G2G}|}{I_{G2C}}$ and the same compared to a C2C path is defined as $\frac{|I_{G2C}-I_{C2C}|}{I_{G2C}}$. Similarly, we define a similar metric for a C2G path.

Table 3 depicts the relative change for different metrics for G2C and C2G paths compared to G2G and C2C paths. The results demonstrate that the capacity, available bandwidth and throughputs of both G2C and C2G paths have a much smaller relative change when compared to C2C paths than G2G paths. Further, the RTTs of G2C and C2G paths are equally close to G2G and C2C paths. Finally, the loss rates of G2C and C2G paths are slightly closer to C2C paths than G2G paths.

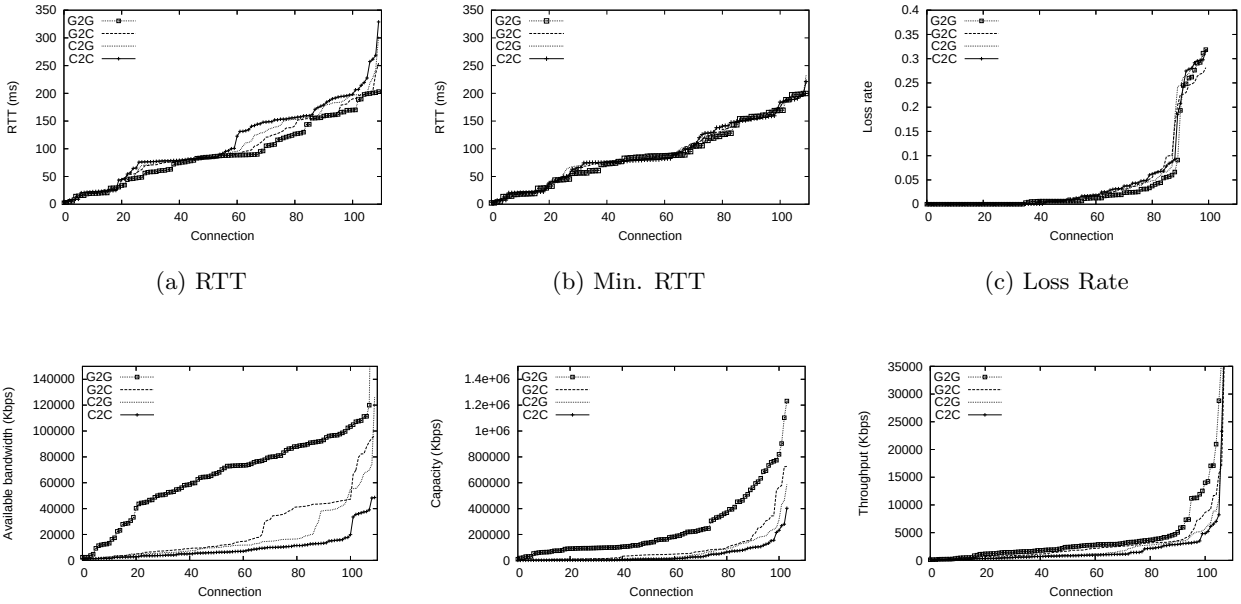
In summary, all traffic flows that contain at least one commercial network connected endpoint are more representative of network conditions in the commercial Internet and can be used to evaluate applications deployed over the testbed. Thus, C2G, G2C and C2C are representative of commercial networks. We now study the impact of mixed testbeds on applications being evaluated.

6. IMPACT ON APPLICATION PERFORMANCE

Sections 4 and 5 show that the representativeness of traffic flow scenarios when an application is deployed over a mixed testbed is affected since when two GREN nodes communicate with each other resulting in overall better performance properties than that experienced by entirely commercial flows. Since the performance properties can affect the application performance in complex ways, in this section, we measure the final impact on the application performance metric via application case studies. The applications are chosen such that they depend on and stress different performance properties.

6.1 Application-Layer Multicast

The first application we evaluate is the popular application-layer multicast. There are several different protocols proposed for providing application-layer multicast. This appli-



(a) RTT (b) Min. RTT (c) Loss Rate

(d) Available Bandwidth (e) Capacity (f) Throughput

Figure 13: Performance properties of Case 2 and Case 3

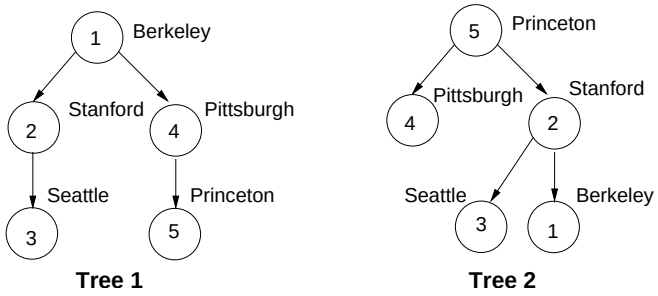


Figure 14: Trees used in application-layer multicast.

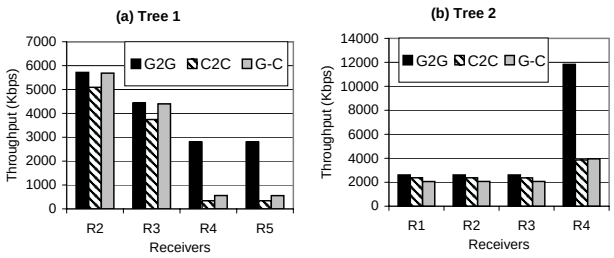


Figure 15: Multicast performance in mixed testbeds.

cation's performance depends on the RTT, path loss and available bandwidth. In this experiment, an application-layer multicast tree is used to transfer a 4 MB file from the source to all the children. Each branch of the tree transfers data reliably using TCP and each branch is independently flow controlled since each node in the tree implements an application layer buffer. The tree is built using a topology-aware protocol similar to that used in NICE [2].

We make two different topologies, each containing five nodes at five different locations as shown in Figure 14. Both trees consist of nodes at Berkeley, Stanford, Seattle, Princeton and Pittsburgh. Tree 1 has a source at Berkeley and Tree 2 has a source at Princeton and thus results in two different topologies. Each location is annotated with a unique number. For each tree, we compare the performance where all nodes for all locations belong to C with that where all nodes in those locations belong to G. Thus for each tree we have a G2G scenario and a C2C scenario.

6.1.1 Results

We first compare the throughput obtained from two different trees shown in Figure 14. For tree 1 we found that the average throughput of receivers in the G2G scenario was 3935.4 Kbps while the average throughput in the C2C scenario was 2379.5 Kbps, a decrease of 40%. Similarly, for tree 2 the average throughput of receivers in the G2G scenario was 4916.7 Kbps while the average throughput in the C2C scenario was 2742.5 Kbps, a decrease of 44.2%.

Figure 15 shows that throughput achieved by each receiver for both trees in C2C and G2G scenarios. In Tree 1, notice that each receiver (2,3,4 and 5) receives the file with lower throughput when the nodes belong to commercial networks compared to when they belong to GREN. Specifically the link from Berkeley to Pittsburgh is particularly worse which then affects the throughput at Princeton. We verified using `pathneck` that the throughput reduction between these nodes in the C2C scenario was in fact due to a bottleneck in the core and not at the access link of the commercial nodes. This can be verified by seeing that in tree 2, the commercial node at Pittsburgh received a throughput of almost 4Mbps. In contrast, the link from Berkeley to Pittsburgh was able to transfer at 3Mbps in the G2G scenario.

In Tree 2, we find that the throughputs are actually comparable in G2G and C2C scenarios for receivers 1, 2 and 3. However, we again see a large gap between the throughput achieved for receiver 4 on the G2G link between Princeton and Pittsburgh (12 Mbps) versus the C2C link between

Conn.	Source	Dest	CP1	CP2	CP3
1	ATT	nbgisp	I/0.872	A/0.599	A/0.070
2	ATT	IRL Berk.	I/0.268	A/0.034	I/0.135
3	ATT	NEC Labs	A/0.017	I/0.009	A/0.006
4	ATT	HP Labs	I/0.790	I/0.296	A/0.028
5	ATT	IRL. Pitt.	A/0.751	I/0.026	I/0.011
6	ATT	atcorp	I/0.193	A/0.148	I/0.012
7	ATT	IRL Seatt.	I/0.164	A/0.043	I/0.016
8	ATT	IRL Camb.	A/0.198	A/0.119	A/0.038
9	ATT	TPG Warsaw	A/0.834	I/0.099	I/0.016
10	ATT	PL-AMST	A/0.206	A/0.049	I/0.035
11	nbgisp	ATT	A/0.027	I/0.024	I/0.019
12	nbgisp	IRL Berk.	I/0.008	I/0.004	I/0.002
13	nbgisp	NEC Labs	I/0.085	A/0.005	I/0.001
14	nbgisp	HP Labs	I/0.026	I/0.011	A/0.002
15	nbgisp	IRL. Pitt.	I/0.195	A/0.025	I/0.007
16	nbgisp	atcorp	I/0.017	I/0.017	A/0.014
17	nbgisp	IRL Seatt.	A/0.009	A/0.005	I/0.003
18	nbgisp	IRL Camb.	I/0.317	I/0.173	I/0.036
19	nbgisp	TPG Warsaw	I/0.097	I/0.053	I/0.010
20	nbgisp	PL-AMST	I/0.003	A/0.002	A/0.001
21	IRL Berk.	ATT	I/0.233	I/0.221	I/0.043
22	IRL Berk.	nbgisp	I/0.098	A/0.061	I/0.020
23	IRL Berk.	NEC Labs	I/0.494	I/0.131	I/0.020
24	IRL Berk.	HP Labs	I/0.065	A/0.040	I/0.005
25	IRL Berk.	IRL. Pitt.	A/0.383	I/0.058	I/0.049
26	IRL Berk.	atcorp	I/0.260	I/0.539	A/0.319
27	IRL Berk.	IRL Seatt.	I/0.063	I/0.037	I/0.014
28	IRL Berk.	IRL Camb.	I/0.672	I/0.398	-/-
29	IRL Berk.	TPG Warsaw	I/0.531	I/0.510	A/0.441
30	IRL Berk.	PL-AMST	A/0.668	I/0.430	-/-
31	NEC Labs	ATT	A/0.511	A/0.017	-/-
32	NEC Labs	nbgisp	A/0.453	I/0.795	A/0.200
33	NEC Labs	IRL Berk.	I/0.934	A/0.435	I/0.020
34	NEC Labs	HP Labs	I/0.953	A/0.430	-/-
35	NEC Labs	IRL. Pitt.	A/0.410	A/0.541	I/0.017
36	NEC Labs	atcorp	A/0.523	I/0.038	I/0.035
37	NEC Labs	IRL Seatt.	I/0.738	A/0.783	I/0.022
38	NEC Labs	IRL Camb.	A/0.495	A/0.559	I/0.034
39	NEC Labs	TPG Warsaw	A/0.914	A/0.505	I/0.034
40	NEC Labs	PL-AMST	A/0.710	I/0.503	-/-
41	HP Labs	ATT	I/0.013	I/0.609	I/0.009
42	HP Labs	nbgisp	I/0.128	A/0.105	A/0.008
43	HP Labs	IRL Berk.	I/0.010	I/0.008	A/0.006
44	HP Labs	NEC Labs	I/0.710	I/0.006	-/-
45	HP Labs	IRL. Pitt.	A/0.132	I/0.004	I/0.023
46	HP Labs	atcorp	I/0.442	I/0.280	I/0.012
47	HP Labs	IRL Seatt.	A/0.051	I/0.013	I/0.013
48	HP Labs	IRL Camb.	I/0.092	I/0.885	A/0.026
49	HP Labs	TPG Warsaw	I/0.746	A/0.515	-/-
50	HP Labs	PL-AMST	I/0.022	I/0.008	-/-
51	IRL. Pitt.	ATT	I/0.012	I/0.001	A/0.001
52	IRL. Pitt.	nbgisp	I/0.067	I/0.048	A/0.038
53	IRL. Pitt.	IRL Berk.	I/0.257	I/0.010	A/0.000
54	IRL. Pitt.	NEC Labs	A/0.011	I/0.003	I/0.003
55	IRL. Pitt.	HP Labs	I/0.032	I/0.002	I/0.001
56	IRL. Pitt.	atcorp	A/0.045	I/0.028	A/0.010
57	IRL. Pitt.	IRL Seatt.	I/0.752	I/0.019	A/0.000
58	IRL. Pitt.	IRL Camb.	I/0.022	I/0.056	I/0.016
59	IRL. Pitt.	TPG Warsaw	I/0.068	A/0.013	I/0.001
60	IRL. Pitt.	PL-AMST	A/0.016	I/0.010	I/0.003
61	IRL Seatt.	ATT	I/0.005	A/0.005	I/0.004
62	IRL Seatt.	nbgisp	I/0.006	I/0.001	I/0.001
63	IRL Seatt.	IRL Berk.	I/0.020	A/0.008	-/-
64	IRL Seatt.	NEC Labs	I/0.174	I/0.039	I/0.036
65	IRL Seatt.	HP Labs	I/0.051	A/0.008	I/0.000
66	IRL Seatt.	IRL. Pitt.	I/0.198	A/0.034	I/0.007
67	IRL Seatt.	atcorp	I/0.045	A/0.008	I/0.007
68	IRL Seatt.	IRL Camb.	A/0.359	I/0.049	I/0.039
69	IRL Seatt.	TPG Warsaw	A/0.143	I/0.091	I/0.044
70	IRL Seatt.	PL-AMST	A/0.006	A/0.001	-/-
71	IRL Camb.	ATT	A/0.043	I/0.040	I/0.037
72	IRL Camb.	nbgisp	I/0.539	I/0.017	I/0.015
73	IRL Camb.	IRL Berk.	I/0.367	A/0.196	-/-
74	IRL Camb.	NEC Labs	I/0.554	A/0.051	-/-
75	IRL Camb.	HP Labs	I/0.088	I/0.445	I/0.682
76	IRL Camb.	IRL. Pitt.	I/0.148	-/-	-/-
77	IRL Camb.	atcorp	I/0.172	I/0.007	I/0.005
78	IRL Camb.	IRL Seatt.	I/0.645	I/0.305	I/0.010
79	IRL Camb.	TPG Warsaw	A/0.714	I/0.999	I/0.729
80	IRL Camb.	PL-AMST	A/0.666	I/0.195	A/0.073
81	TPG Warsaw	ATT	I/0.005	A/0.004	I/0.000
82	TPG Warsaw	nbgisp	A/0.005	I/0.002	I/0.001
83	TPG Warsaw	IRL Berk.	I/0.034	I/0.003	I/0.003
84	TPG Warsaw	NEC Labs	I/0.035	I/0.005	A/0.002
85	TPG Warsaw	HP Labs	I/0.013	I/0.005	I/0.002
86	TPG Warsaw	IRL. Pitt.	I/0.233	A/0.175	I/0.140
87	TPG Warsaw	IRL Seatt.	A/0.014	I/0.013	A/0.002
88	TPG Warsaw	IRL Camb.	A/0.011	I/0.003	A/0.002
89	TPG Warsaw	PL-AMST	A/0.002	I/0.000	I/0.000
90	PL-AMST	IRL Pitt.	A/0.866	I/0.123	-/-
91	PL-AMST	IRL Camb.	I/0.088	-/-	-/-

Table 2: Bottleneck locations in C2C paths. "-/-" denotes that the tool failed to locate any choke points. Paths from atcorp and a few other paths are not depicted due to failure in detecting bottlenecks.

Paths	RTT	Loss	Cap.	Av. BW	Thrput.
G2C-G2G	13.7	77.01	843.2	793.1	267.5
G2C-C2C	13.0	67.66	123.3	68.7	49.2
C2G-G2G	15.1	139.8	1359.1	704.3	292.2
C2G-C2C	14.1	97.6	307.0	45.3	76.3

Table 3: Summary of comparisons using the relative change metric.

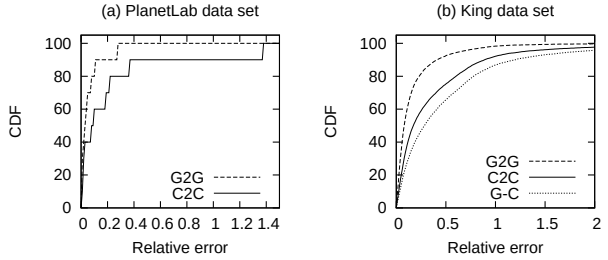


Figure 16: GNP performance in mixed testbeds.

Princeton and Pittsburgh that achieves 4 Mbps.

In conclusion, application performance can significantly vary depending on how the testbed is utilized. The research network can significantly affect the measured performance. If a multicast tree was indiscriminately constructed with nodes belonging to both G and C, the G2G branches would typically have higher throughput than a C2C link of similar RTT. However if the tree is chosen only to have G2C or C2G links, the performance is similar to C2C. For example in Tree 1, we made the source node 1 a GREN node and chose the remaining nodes in the given locations such that all links are either G2C or C2G[3]. Figure 15(a) shows that this G-C case is similar to C2C. Similarly, for Tree 2 we chose the source node 5 to be a commercial node and again chose the remaining nodes maintaining G2C/C2G links and the performance is similar to C2C (Figure 15(b)).

6.2 Network Distance Prediction

Network distance prediction is an important network service that allows the prediction of N^2 RTTs between N nodes using only $O(N)$ measurements. The performance, i.e., accuracy, of such prediction services depend on the RTT, connectivity, peering policies, etc. Here we study the impact of a mixed testbed on network distance prediction using GNP [11], one of several systems proposed to perform network distance prediction. GNP uses a fixed set of landmarks which first compute their coordinates. Each end host then derives its coordinates by minimizing the error between measured and estimated distances to landmarks.

6.2.1 Results

We used two real Internet latency datasets to evaluate GNP (PlanetLab and "King" [34]). Each data set is essentially an all-pair RTT matrix between a set of nodes. For each data set we filtered out the hosts into GREN and commercial nodes and extracted smaller all-pair RTT matrices. The PlanetLab data set is based on all-pair RTT measurements among 8 GREN nodes and among 8 commercial nodes from Section 5. The King data set originally had RTT measurements between 1,953 nodes. We found 350 GREN nodes were part of this set so we also extracted

[3]The tree is 1∈G, 2∈C, 4∈C, 3∈G, 5∈G.

out 350 commercial nodes randomly from the trace. Since the GNP algorithm requires a completely filled matrix, we ran the Bron-Kerbosch clique generation algorithm [5] and found a complete all-pair RTT matrix among 164 GREN nodes and among 172 commercial nodes. We ran the publicly available GNP code [30] on these traces to find the relative error in the predicted and measured distances, i.e., $\frac{|d_{predicted}-d_{measured}|}{d_{measured}}$ with 15 randomly chosen GREN landmarks for the GREN nodes and commercial landmarks for the commercial nodes.

Figure 16 shows that for both the smaller PlanetLab data set and the larger King data set, running GNP on a G2G testbed has much lower relative error than on a C2C testbed. Thus the performance of network distance prediction depends on the nature of nodes selected in the testbed. We believe the G2G data set has lower errors because the G2G network has fewer routing inefficiencies and its RTTs are closely related to the node placement rather than queuing and peering policies. The C2C testbed can have many such variable quantities that make it difficult to accurately predict network distances. Finally, we used a mixed trace with 210 commercial nodes and 140 GREN nodes selected randomly from the larger King data set localized by 15 commercial landmark nodes. The error distribution in this case is more representative and similar to the C2C scenario since a good number of G2C/C2G links were involved. The error is slightly worse than C2C since localizing G2G links using commercial landmarks can cause triangle inequalities.

In summary, G2G links are not representative of C2C links, both with respect to performance properties and application metrics. On the other hand, G2C and C2G better represent C2C links. Thus, when a mixed testbed is used to evaluate applications, G2G links should be avoided. In the next section, we advocate a technique to leverage the existing mixed testbeds while remedying the G2G links.

7. RECOMMENDATIONS ON USING MIXED TESTBEDS

Based on the observations in the previous sections, we now propose a *node partitioning* principle that minimizes the impact of mixed testbeds on application performance while leveraging the available GREN nodes simultaneously. Node partitioning takes a set of N nodes from a testbed and *partitions* them into sets C and G where C contains nodes with only commercial connectivity and G contains nodes that can route over GREN. A NP graph is then built such that an edge between nodes i and j is inserted if and only if (1) $i \in G$ and $j \in C$ or (2) $i \in C$ and $j \in G$ or (3) $i \in C$ and $j \in C$. This graph is then used by applications running on the testbed nodes.

7.1 Using Node Partitioning

We have shown that Cases 2 and 3 result in traffic predominantly flowing over commercial networks and that the performance properties observed are similar to Case 4, thereby reducing the impact on application performance. Based on this insight, a simple approach to improve the representativeness of application performance is to integrate node partitioning into distributed applications. For example, consider overlay multicast applications. These typically use neighbor beaconing to form multicast trees. Such applications can then be modified so that their neighbor view is restricted as per the rules (1), (2) and (3) above. In this case, the entire multicast tree constructed will be such that traffic flows primarily over commercial networks and is thereby representative.

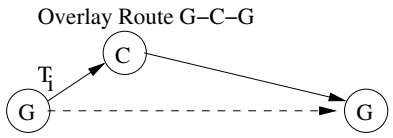


Figure 17: Illustration of NP.

A limitation with the above approach is that not all distributed applications can tolerate such disconnection. Consider a DHT application on top of Pastry [20]. Assume when all the overlay members are considered, the closest node to item K is D. When S hashes K, if D is in S's leafset but $S \rightarrow D$ does not satisfy the rules above, data item K cannot be inserted. This will lead to inconsistencies in the DHT. Thus we need a method that improves representativeness transparently to the applications. To this end, we propose a toolkit, *NP*.

NP leverages the techniques from overlay routing [1, 7] while partitioning nodes into those with GREN connectivity. *NP* transparently captures packets from applications on particular *slices* based on pre-specified port numbers. If the host node is a commercial node, then packets are left unchanged. If the host node is a GREN node, the destination IP is extracted from the packets and looked up in a database to identify whether it is a GREN host. If the destination is a GREN host, *NP* resorts to overlay routing (as shown in Figure 17 and routes all further packets to this destination via an appropriately chosen intermediate commercial host using an overlay route G-C-G. This can be easily done using the same technique used to repair routes in SOSR. The intermediate commercial host is chosen such that it lies en-route to the GREN destination. For example, the commercial host chosen could be a node closest in network distance (i.e., minimizes T_i) to the source/destination GREN node so as to minimize detours.

Applicability: Using *NP* should be a choice made by the application designer when using the testbed and not a general purpose service. For example, research that performs network measurements should not use *NP* since the overlay routing will not reflect the intent of the measurements and will provide erroneous results.

NP is also limited by the number of commercial nodes available in the testbed. However, we argue that it may be useful just to increase the number of machines at commercial sites[4]. This allows for a limited number of commercial sites to be used for a large number of connections. However, *NP* will not be useful for sender/receiver GREN node pairs that have no commercial site close (in network distance) to either the sender or the receiver.

7.2 Improving G2G Path Diversity

While *NP* eliminates G2G links, another approach to improving representativeness in a mixed testbed is to include G2G links that come from diverse GREN ASes [3]. From our current measurements in PlanetLab, we found that out of a total of 1322 GREN ASes, our all pair traceroutes (between 155 GREN nodes) traversed only 182 unique ASes. Further,

[4]It is typically easier (cheaper and convenient) to add more machines at a current site than add an entirely new site.

out of all possible GREN AS → GREN AS links, only 901 unique links were visited. Thus there is a need and scope to improve the diversity in PlanetLab. Also, when such diverse GREN ASes become part of the testbed, it is essential to revisit the performance properties of G2G paths (which may then have increased diversity) and compare them to C2C paths. Note that even if there is improvement in network diversity within GREN, an inherent limitation is that traffic in GREN is much lesser than in commercial ISP networks.

As part of our future work, we are investigating if the distribution of performance properties of C2C links can be emulated over G2G links in order to leverage G2G links while improving the representativeness of wide-area experiments.

8. CONCLUSIONS

In this paper, we assessed the potential impact of currently used wide area network testbeds on the measured performance of distributed systems and network services that are widely deployed and prototyped on such testbeds. We found that the fact that a large fraction of nodes are connected to research networks can impact the representativeness of testbed evaluation results. Specifically, we found that G2G paths have significantly different topological and performance properties (available bandwidth, throughput, etc.) than commercial network paths over which the distributed systems and network services are to be deployed. Encouragingly, we found that traffic that traverses from a G to a C node and vice-versa is more representative of commercial networks since such paths primarily traverse commercial networks and were found to have similar performance properties to C-C paths. We demonstrated how the testbed connectivity can impact application evaluation results and proposed a technique that can be leveraged by applications to improve the representativeness of their evaluation results while maximally leveraging the existing infrastructure.

Our future work includes using routing data to assess the routing stability of mixed GREN and commercial paths and use it as an additional metric to measure the testbed representativeness, as well as implementing the *NP* technique and evaluating its performance in practice.

Acknowledgments

We would like to thank the anonymous reviewers for their helpful comments. This work was supported in part by the National Science Foundation under grants CNS-0430204 and CAREER award CCF-0238379.

9. REFERENCES

[1] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris. Resilient Overlay Networks. In *Proc. of ACM SOSP*, 2001.
[2] S. Banerjee, B. Bhattacharjee, and C. Kommareddy. Scalable Application Layer Multicast. In *Proc. of ACM SIGCOMM*, 2002.
[3] S. Banerjee, T. G. Griffin, and M. Pias. The Interdomain Connectivity of PlanetLab Nodes. In *Proc. of PAM*, 2004.
[4] A. Bavier, N. Feamster, M. Huang, J. Rexford, and L. Peterson. In VINI Veritas: Realistic and Controlled Network Experimentation. In *Proc. of SIGCOMM*, 2006.
[5] C. Bron and J. Kerbosch. Algorithm 457: Finding All Cliques of an Undirected Graph. *Commun. ACM*, 16(9), 1973.
[6] C. Dovrolis, P. Ramanathan, and D. Moore. What Do Packet Dispersion Techniques Measure? In *Proc. of IEEE Infocom*, 2001.
[7] K. P. Gummadi, H. Madhyastha, S. D. Gribble, H. M. Levy, and D. J. Wetherall. Improving the Reliability of Internet Paths with One-hop Source Routing. In *Proc. of OSDI*, 2004.
[8] N. Hu, L. E. Li, Z. M. Mao, P. Steenkiste, and J. Wang. Locating internet bottlenecks: algorithms, measurements, and implications. In *Proc. of SIGCOMM*, 2004.
[9] Y. hua Chu, S. G. Rao, and H. Zhang. A Case for End System Multicast. In *Proc. of ACM SIGMETRICS*, 2000.
[10] R. Mahajan, N. Spring, D. Wetherall, and T. Anderson. User-level internet path diagnosis. In *Proc. of SOSP*, 2003.
[11] T. S. E. Ng and H. Zhang. Predicting Internet Network Distance with Coordinates-Based Approaches. In *Proceedings of IEEE INFOCOM*, June 2002.
[12] U. of Oregon Route Views Archive Project. http://www.routeviews.org.
[13] K. Park, V. S. Pai, L. Peterson, and Z. Wang. CoDNS: Improving DNS Performance and Reliability via Cooperative Lookups. In *Proc. of OSDI*, 2004.
[14] V. Paxson. Strategies for sound internet measurement. In *Proc. of IMC*, 2004.
[15] L. Peterson, T. Anderson, D. Culler, and T. Roscoe. A Blueprint for Introducing Disruptive Technology Into the Internet. In *Proc. of ACM HotNets*, 2002.
[16] PlanetLab. http://www.planet-lab.org.
[17] S. Rhea, B. Godfrey, B. Karp, J. Kubiatowicz, S. Ratnasamy, S. Shenker, I. Stoica, and H. Yu. OpenDHT: A Public DHT Service and Its Uses. In *Proc. of SIGCOMM*, 2005.
[18] RipeNCC: Routing Information Service Raw Data. http://abcoude.ripe.net/ris/rawdata/.
[19] RON. http://nms.csail.mit.edu/ron/sites/.
[20] A. Rowstron and P. Druschel. Pastry: Scalable, Distributed Object Location and Routing for Large-Scale Peer-to-peer Systems. In *Proc. of ACM/IFIP/USENIX Middleware*, November 2001.
[21] A. Rowstron and P. Druschel. Storage management and caching in PAST, a large-scale, persistent peer-to-peer storage utility. In *Proc. of SOSP*, 2001.
[22] J. Sommers, P. Barford, N. Duffield, and A. Ron. Improving accuracy in end-to-end packet loss measurement. In *Proc. of SIGCOMM*, 2005.
[23] N. Spring, R. Mahajan, and T. Anderson. Quantifying the causes of internet path inflation. In *Proc. of SIGCOMM*, 2003.
[24] J. Strauss, D. Katabi, and F. Kaashoek. A measurement study of available bandwidth estimation tools. In *Proc. of IMC*, 2003.
[25] L. Subramanian, I. Stoica, H. Balakrishnan, and R. Katz. OverQoS: An Overlay Based Architecture for Enhancing Internet QoS. In *Proc. of USENIX NSDI*, 2004.
[26] L. Wang, K. Park, R. Pang, V. S. Pai, and L. Peterson. Reliability and security in the codeen content distribution network. In *Proc. of USENIX ATC*, 2004.
[27] B. White, J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M. Newbold, M. Hibler, C. Barb, and A. Joglekar. An Integrated Experimental Environment for Distributed Systems and Networks. In *Proc. of OSDI*, 2002.
[28] P. Yalagandula, P. Sharma, S. Banerjee, S.-J.Lee, and S. Basu. S3: A Scalable Sensing Service for Monitoring Large Networked Systems. In *Proc. of the Workshop on Internet Network Management*, 2006.
[29] R. Zhang and Y. C. Hu. Assisted Peer-to-Peer Search with Partial Indexing. In *Proc. of IEEE INFOCOM*, 2005.
[30] GNP Homepage. http://www.cs.rice.edu/ Eugeneng/research/gnp/.
[31] Iperf. http://dast.nlanr.net/Projects/Iperf/.
[32] PlanetLab IPerf. http://www.planet-lab.org/logs/iperf/.
[33] S3: Scalable Sensing Service. http://networking.hpl.hp.com/s-cube/.
[34] The P2PSim Project. http://pdos.csail.mit.edu/p2psim/.
[35] The Gnutella protocol specification, 2000. http://dss.clip2.com/GnutellaProtocol04.pdf.