On the Use of Anycast in DNS

Sandeep Sarat
Johns Hopkins University
sarat@cs.jhu.edu

Vasileios Pappas
UCLA
vpappas@cs.ucla.edu

Andreas Terzis
Johns Hopkins University
terzis@cs.jhu.edu

Abstract

We present the initial results from our evaluation study on the performance implications of anycast in DNS, using four anycast servers deployed at top-level DNS zones. Our results show that 15% to 55% of the queries sent to an anycast group, are answered by the topologically closest server and at least 10% of the queries experience an additional delay in the order of 100ms. While increased availability is one of the supposed advantages of anycast, we found that outages can last up to multiple minutes, mainly due to slow BGP convergence. On the other hand, the number of outages observed was fairly small, suggesting that anycast provides a generally stable service.

Categories and Subject Descriptors: C.2.3[**Computer Communications and Networks**]: Network Operations - *network management*.
General Terms: Measurement

1. INTRODUCTION

Anycast, where client packets are directed towards the server in the anycast group closest to the client, is widely deployed in DNS today [4]. While not visible to end-users, the IP addresses of many top level DNS nameservers actually correspond to anycast groups. Client requests sent to these addresses are delivered by the Internet routing infrastructure to the closest replica in the corresponding anycast group. DNS operators have deployed anycast for a number of reasons: reduced query latency, increased reliability and resiliency to DDoS attacks. While it is generally agreed that the deployment of anycast in DNS has been a positive step, no studies exist that evaluate the performance improvement offered by anycast. This work is an initial attempt at such a study. Specifically, we are interested in answering the following questions: (1) Does anycast selection indeed lead clients to the closest DNS server? and (2) Do zones deploying anycast experience smaller number of outages and if yes, how long are these outages? The technical report [1] discusses these and other findings in greater detail.

To answer these questions, we performed a large scale measurement study using clients deployed over the PlanetLab testbed to measure the performance characteristics of four anycast servers, deployed at three different top-level DNS zones, and compared it to legacy DNS servers. The monitored anycast servers represent two different points on the anycast design space and we use them to compare the effects of these design choices. We also compared these zones against a hypothetical zone with the same number of nameservers but where all the nameservers are individually addressable. By doing so, we can directly compare anycast to the traditional zone configuration guidelines [3].

SIGMETRICS'05, June 6–10, 2005, Banff, Alberta, Canada.
ACM 1-59593-022-1/05/0006.

2. MEASUREMENT METHODOLOGY

We used the following four types of servers in our measurements, each representative of a different deployment scenario: **(1)** A server with one or more instances in a single geographic location: While this case does not use anycast, we use it as a base case to explore the potential performance improvements of anycast. We chose the B-root nameserver (192.228.79.201) as a representative of this category. **(2)** A server using a single anycast address for all its instances, deployed in multiple locations, with only a subset of the instances being globally visible, and the rest being advertised only locally. Since this anycast configuration is widely used, we chose two different examples, the F-root nameserver (192.5.5.241) and the K-root nameserver (193.0.14.129). Choosing two examples enables us to investigate the effects of the number and location of anycast group members on performance **(3)** A server using a single anycast address for all its instances, deployed in multiple locations, with all the instances being globally visible. We used UltraDNS servers, which are authoritative for the .org and .info top level domains, as the representative examples of this category. UltraDNS servers are members of two anycast groups TLD1 (204.74.112.1) and TLD2 (204.74.113.1). The benefit of using multiple anycast addresses is that in the event of a network outage affecting one of the anycast addresses, clients can query the other address for resolving names. **(4)** A collection of multiple geographically distributed servers, each individually accessible via unicast. To emulate this scenario, clients send requests to the unicast addresses of the F-root group members. Each client maintains an ordered list of servers based on their latency. This list is updated every hour. During each hour all DNS queries are directed to the closest server according to this ranking. If the current server becomes unavailable, the client tries the servers from its server list until it receives a response. We use this scenario to compare anycast to unicast when the number and location of name servers is the same.

Because DNS clients have no control of where their queries are directed, we need clients in multiple locations to cover all the servers in an anycast group. For this reason, we used the PlanetLab [7] testbed for our measurements.

We ran a script on every PlanetLab node to send periodic DNS queries (the query interval is randomly selected between 25 and 35 seconds) to each of the DNS zones mentioned earlier. To ensure that the query time includes only a single round trip, we send queries that can always be answered locally by the server receiving the request. Our script records the query latency and unicast name corresponding to the server in the anycast group answering the query. To aid debugging, anycast servers return the unicast name of the server answering a special anycast request. For example, to obtain the unicast address of the server from the F-root anycast group answering a query, we query for the `hostname.bind txt` record in the `chaos` class [2]. Below is an example of such a query and the server response.

```
$dig +short +norec @192.5.5.241 hostname.bind
chaos txt
"sfo2a.f.root-servers.org"
```

The results presented in this paper are based on data collected during the period from September 19, 2004 to October 8th, 2004.

3. EVALUATION

This section examines (1) the percentage of clients not reaching the replica server that is closest to them and the additional delay incurred (2) the availability of the monitored zones.

3.1 Effectiveness of Localization

Anycast server selection depends on the path selected by BGP. The unicast addresses of the DNS servers are selected from different address ranges and therefore the path from a client to the anycast address might be different from the path to the unicast address of the server. For these reasons a direct comparison does not yield accurate results. Hence, to measure the effectiveness of localization for each of the anycast zones, we compare the geographic distances between the client and the geographically closest server to the geographical distance between the client and the actual anycast server contacted. While it is known that Internet paths are longer than the direct geographic path connecting two end-points [6, 8], we assume that all paths exhibit the same *path inflation* factor. Based on this assumption, we can directly compare geographic distances to determine whether the best Internet path is selected for each client.

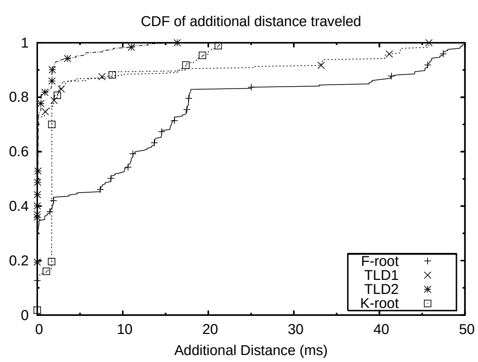

Figure 1: Additional distance over the optimal traveled by anycast queries to contact their F-root, K-root, TLD1 and TLD2 server.

Figure 1 presents the cumulative distribution of the additional distance across all PlanetLab nodes for each zone. We observe that only 27% of all the anycast requests are sent to the nearest F-root server while 15% of the anycast requests are sent to the nearest K-root server. Approximately 53% requests are served by the nearest TLD1 server while for TLD2 the percentage is about 45%. Differentiation of global/local clusters leads to the smaller proximity metric for F-root and K-root, though F-root performs marginally better due to a greater number of clusters. TLD1/TLD2 perform the best as clients have visibility to a greater number of BGP routes to the clusters and can thus can make a better choice.

3.2 Availability

We use the term "outage" to indicate a window of time when a client is unsuccessful in retrieving a record from the anycast server servicing it. For all the measured zones, the average number of outages per node per hour is extremely low ($\leq 0.4\%$). Nameservers having clusters geographically distributed had greater availability.

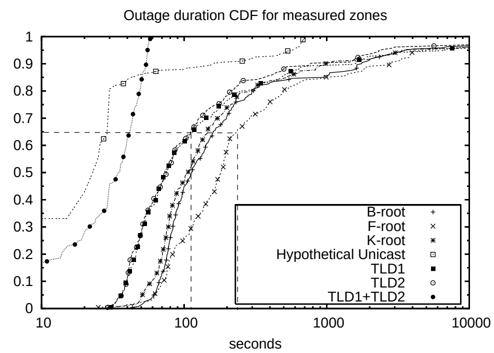

Figure 2: CDF of outage duration

Fig. 2 plots the outage duration CDFs for all the nameservers. The existing anycast services have considerable overlap and exhibit similar behavior. This reveals an interesting fact regarding anycast. Since anycast relies on Internet routing, once an outage has occurred the recovery time of anycast is governed by the recovery time of the network routing fabric. In fact, about 35% of the outages span a duration higher than 180 seconds. This is a direct consequence of the results presented by Labovitz et.al regarding delayed network convergence [5]. The outage recovery time is independent of the anycasting scheme used. What different anycast setups try to achieve is to reduce the number of outages.

4. SUMMARY

We found that in general the deployment of anycast decreases the average query latency. On the other hand, the majority of clients (55%-85%) are not directed towards the closest anycast server. Making servers globally visible improves the proximity metric of anycast selection. We also observed that anycast provides a fairly stable service, but due to its reliance on Internet routing, outages can last up to multiple minutes due to slow BGP convergence.

5. REFERENCES

[1] S.Sarat V. Pappas, and A.Terzis, On the Use of Anycast in DNS, *HiNRG, Johns Hopkins University Technical Report,Dec. 2004*.

[2] J.Abley, A Software Approach to Distributing Requests for DNS Service Using GNU Zebra, ISC BIND 9, and FreeBSD, *Proceedings of USENIX 2004 Annual Technical Conference, FREENIX Track, 2004*

[3] R.Elz, R.Bush, S.Bradner and M.Patton, Selection and Operation of Secondary DNS Servers, July 1997.

[4] T.Hardie, Distributing Authoritative Name Servers via Shared Unicast Addresses, *RFC 3258*, Apr.2002.

[5] C.Labovitz, A.Ahuja, A.Bose and F.Jahanian, Delayed Internet Routing Convergence, *Proceedings of ACM SIGCOMM 2000*, 2000.

[6] L. Gao and F. Wang, The extent of AS path inflation by routing policies, *Proceedings of Global Internet Symposium, 2002*.

[7] Intel Research, Planet Lab, *http://www.planet-lab.org/*.

[8] N.Spring, R.Mahajan and T.Anderson, Quantifying the Causes of Path Inflation, *In Proceedings of ACM SIGCOMM*, Aug. 2003.