

Diversity in DNS performance Measures

Richard Liston, Sridhar Srinivasan, Ellen Zegura

Review

This paper gives different measures to find out the geographical, topological influences on the DNS performance in wide area networks. This paper is of importance because the data collected is the latest and from many locations around the world. The measurements done in this paper would help the future engineering studies of DNS. Not only they observe the performance of the root servers, but also conclude about the distribution of the gTLD serves that if they are placed more on an equitable choice, they would show a conceivable change in the performance DNS requests.

Some significant features of this paper :

- They verified with their measurements that Root DNS is only involved for 7% of the requests, which is a claim, made by a different paper.
- Currently TTLs do not depend on the location of the client.
- Tools like the named, dig and gethostbyname provided with BIND are used.
- They observed that invalid domain names requested also depend on the time of the year. Requests from the month of Jan 2002 has 0.6% less invalid domain name requests form the month of Mar / Apr the same year.
- For the non-cached domain names the performance is significantly effected by the servers other than the root and gTLD.

Possible improvements :

DNS requests considered in the paper are from the local name server rather than the actual client. They could have extended their paper to study the performance of the requests obtained from the clients in different locations also.

The test data only consists of the non-cached domain names but could have extended their work to verify the conclusions they made from the paper.

Clients can directly contact the gTLD servers, so there could be some of the requests that were neglected from the performance study.

Their conclusion that gTLD servers are location sensitive is evident from the Fig.2 in the paper. Since the gTLD servers are not equally distributed and at the same time, there are few links going out from some of the countries like India, the conclusion drawn seems quite evident.

The seed point of the crawler to start, for collecting the domain names is strange and not uniform, some of the seeds were famous sites in USA and others were not so famous sites from other parts of the world. Considering a specific case, the seed site from India is strange, not so popular government site from India. Not sure what type of links they might have collected from such a site. They could have based their selection on the page rank of the sites from different countries that gives some basis for their selection and they could have applied some statistical models for their selection and compared the performance applying various models. The paper does not explain the basis of the selection of the seed points for the crawler. Though it is a not an important issue for the results of the experiment. It could have influence on the measurements varying for the different types of sites requested.

In the section III part A, they say that some collection sites were found broken while they were verifying the data with extra information and due to that they did not include those datasets. At the same time they say this extra information is only available for some of the sites, so there is a possibility that the data presented from some of the sites could be the broken sites for which the extra data is not available. This case was not considered.

Coming back to the seed point of the domain collection countries (refer to TABLE 1) and the list of countries that they have run their collection software tools is quite not related. For instance consider the seed page of the country India; they start with a government web site which is related to a state called Himachal Pradesh and observe the countries at which the data collection software is run. Just wondering why would a user from the country like Argentina would like to go this web site. There are two points I would like to make in this matter. First, the requests could be regional / country specific and secondly, since there is no relation between the domain names and the data collection countries their collected set of data might not truly reflect the user behavior on the net. The concern in this paper is about the DNS performance through out the world but simulated performance measures could vary if the actual regional popular sites and regional rarely visited sites could be collected by some heuristics.

For calculation of Mean Response Time and others, they have considered equal representation of the datasets collected. But the results could have been more representative of the internet if the data is given weightage according to the percentage usage in the specific region.

Future Work:

The authors have given a very useful future work, encouraging others to supplement the work from other countries and other list of ponder able research topics.