Independent Comment to

"Name Collision in the DNS - A study of the likelihood and potential consequences of collision between new public gTLD labels and existing private uses of the same strings"

prepared by

eco - Verband der deutschen Internetwirtschaft e.V.

Motivation

eco Verband der deutschen Internetwirtschaft e.V. is one of the leading Internet Industry Associations, with a highly diverse membership of some 700 member organizations originating from over 60 countries worldwide. These include ISPs, Registries, Registrars and new gTLD applicants.

eco's division dealing with topics related to the domain industry, the eco Names and Numbers Forum has learned of the study prepared by Interisle Consulting Group concerning the deployment of new public gTLD labels and expressed an interest in its content.

During our deliberations it has quickly become clear that the number of requests posted to the root name servers is only partially representative of the number of request issued by the global user base. Today, an end user's device will typically resolve DNS names through inquiry at a name resolver located either within his local CPE device or with his ISP directly. In case of a corporate user there will almost always be at least one local resolver present on site. Typically, these devices will do recursion for their users and utilize result caching as a method to reduce the number of queries required in order to speed up name resolution. However, doing so will significantly distort the number of actual external queries performed vs. the number of DNS names requested.

This effect is expected to be more prominent for popular names as well as 'existing' parts of the DNS tree, which can be cached, vs. sparsely used names or "non existent" domains.

A study of actual request numbers and the relative impact of new strings deployed as gTLD names should therefore in our view not be performed at the root servers, but should consider request strings as close to the source as possible.

Also, as already mentioned in the study, an analysis at source level might provide some insight on the source distribution of request for formerly unused names as well as the possible effects of mitigation techniques to reduce the number of requests.

It was therefore decided to verify the findings of the Interisle study through independent research within our member base.

Methodology

We have asked a number of our ISP member organizations to help conduct such a study, utilizing a given date, and were able to collect roughly 7.985 million “real world” queries of more than 2.000.000 users as a basis for comparative, statistical DNS name usage.

Given the limited time frame to conduct the study, we have focused on providers located in Germany only. A variation of the results compared to the Interisle study based on both geographic and language preferences is therefore expected.

In an additional step we have taken an in-depth look at the source of the existing queries for proposed new gTLD domain names in order to weight the potential security impact posed by an immediate deployment of said gTLDs.

Analysis

The distribution of the top 50 most used TLD strings in all queries analyzed can be found in the following table:

Top 50 User TLD Queries Issued (non provider cached)

German Access Providers	Date	12. Sep 13
	Total Queries	~ 6.8 billion..

Toplevel-Domain	Queries (1000')	Query %	Existing	Proposed	Potential
com	3.398.998	50,43%	Yes		
net	1.120.657	16,63%	Yes		
de	720.930	10,70%	Yes		
org	436.071	6,47%	Yes		
arpa	379.782	5,64%	Yes		
tv	171.134	2,54%	Yes		
info	89.138	1,32%	Yes		
hk	77.485	1,15%	Yes		
nu	51.627	0,77%	Yes		
local	45.133	0,67%			Yes
lan	44.225	0,66%			Yes
ru	11.525	0,17%	Yes		
me	11.173	0,17%	Yes		
eu	10.222	0,15%	Yes		
uk	9.981	0,15%	Yes		
cn	8.430	0,13%	Yes		
pl	7.343	0,11%	Yes		
biz	6.170	0,09%	Yes		
it	6.135	0,09%	Yes		
mars	5.833	0,09%			Invalid
at	4.960	0,07%	Yes		
gov	4.959	0,07%	Yes		
tw	4.936	0,07%	Yes		
fm	4.701	0,07%	Yes		
box	4.356	0,06%		Yes	

us	4.216	0,06%	Yes		
fr	3.931	0,06%	Yes		
corp	3.897	0,06%		Yes	
ph	3.671	0,05%	Yes		
in	3.335	0,05%	Yes		
to	3.302	0,05%	Yes		
ch	2.937	0,04%	Yes		
es	2.841	0,04%	Yes		
nl	2.834	0,04%	Yes		
localdomain	2.833	0,04%	Yes		
global	2.711	0,04%		Yes	
cc	2.646	0,04%	Yes		
io	2.551	0,04%	Yes		
mobi	2.526	0,04%	Yes		
jp	2.487	0,04%	Yes		
kr	2.307	0,03%	Yes		
br	2.049	0,03%	Yes		
dfs	2.025	0,03%			Invalid
ca	1.949	0,03%	Yes		
be	1.907	0,03%	Yes		
mx	1.744	0,03%	Yes		
co	1.702	0,03%	Yes		
int	1.699	0,03%	Yes		
cz	1.644	0,02%	Yes		
All Other	40.006	0,59%			
Total	**6.739.654**	**100,00%**			

The above represents the top 50 strings of a total of roughly 1.400 strings found in the queries analyzed. More specifically, 99% of all queries are represented by 40 TLD strings, while 99.9% are represented by 117 TLD strings and 99.99% by 316 TLD strings.

Roughly 1.2 billion queries received were for local, non-reversed IP address to name resolution requests and were not considered for further analysis.

Findings concerning all queries

As expected, the distribution of requests for gTLD names is similar to the order of queries received at the root, while the order of ccTLD requests is different based on a more specific geographical location.

Most notable is the almost complete absence of “invalid” requests vs. the 23% found in the Interisle study. An excessive number of invalid top level domain strings can indeed be found in the requests processed, however the relative number is so low compared to the total number of requests processed that the invalid names can be found alongside the rest of the ccTLDs, gTLDs and reserved names in the “all other” category (with the notable exception of .mars and .dfs).

Also, the effect of name caching discussed above can prominently be found in the number of requests into the .com and .net domain: 13.5 of 39 billion (34.6%) requests in the Interisle root server study vs. 4.5 of 6.7 billion (67,1%) requests in the numbers analysed here.

It should be noted that the results are in general similar to the figures presented in chapter 4.2 “request stream at intermediate resolvers” of the Interisle study, which were unfortunately not used as the basis of the TLD string analysis.

Proposed new gTLD strings

A closer look at the number of requests found for the proposed new gTLD names result in a vastly different set of numbers as compared to the Interisle study:

Interisle	39.000.000	(all numbers are in 1.000')
eco	6.739.000	

new	Rank	string				
gTLD	(Interisle)	count	Percent	Rank (eco)	string count	Percent
home	1	595.024	1,53%	4	328	0,00%
corp	2	122.794	0,31%	2	3.897	0,06%
site	3	13.013	0,03%	6	4	0,00%
global	4	10.838	0,03%	3	2.711	0,04%
ads	5	7.799	0,02%	n.a.	0	0,00%
iinet	6	7.668	0,02%	n.a.	0	0,00%
group	7	6.505	0,02%	n.a.	0	0,00%
box	8	6.152	0,02%	1	4.356	0,06%
cisco	9	5.231	0,01%	n.a.	0	0,00%
hsbc	10	4.924	0,01%	n.a.	0	0,00%
inc	11	4.622	0,01%	n.a.	0	0,00%
network	12	4.417	0,01%	n.a.	0	0,00%
dev	13	4.344	0,01%	5	4	0,00%

It should be noted that a significant number of names listed could not be found in the request set at all, i.e. .hsbc, .iinet or .ads were not present in any requests received at the resolver sites, while for established names even typographical errors were present. Obviously, these name strings are only popular in other geographical regions – i.e. were the entity in question is active or the string holds another meaning altogether.

Interestingly enough, the top proposed gTLD name string considered - “.home” - could only be found at resolver sites in trivial numbers, while the “.box” proposes name string ranked even higher than the .us, .fr, .es or .nl ccTLD domains – owing to the vendor dominance of one particular brand of CPE devices in Germany.

Only three of the proposed gTLD names make it to the list of top 50 strings queries at all: .box, .corp and .global, composing a total of 10,9 from 6.739 queries or 0,16% of total queries. This figure is roughly 1/10th of the implied number by the root server study for the .home domain alone.

While this does not reduce the security concerns raised for each individual - potentially exploitable - allocation, the total impact to be expected from the delegation of additional name strings at the root seems significantly lower than suggested for even the top proposed names and virtually non-existent for most of the name strings.

Further Work

Following the findings above, an additional analysis has been undertaken to track down the sources of queries for the most prominent non-existent TLD domain names, namely .box, .corp and .global:

Proposed gTLD	requests (1.000')	Individual Strings	Unique Sources
global	2.711	1636	410
corp	3.897	330	291
box	4.356	25	3.705

These findings were in-line with the expectation, namely that .box as a vendor-specific extension in universal use has a broad base of origination while .corp and .global were related mostly to VPN gateway and intranet usage with a highly localized point of origination.

For .corp, over half of the actual requests originated from only 25 sites at two participating ISPs. In a quick mitigation trial, 8 of these 25 sites could already be resolved through ISP intervention in less than 24h after discovery – in any case a good example of DNS cleanup.

Conclusion

It can be demonstrated that the usage of proposed new gTLD strings is strongly related both to the geographic location as well as vendor dominance in the individual market, making an assessment of the global impact of individual name delegation close to impossible - even if data could be retrieved from all root servers it will with a high probability yield differing results based on the geographical location of the individual server. In any case, the results presented by the Interisle study are only a snapshot and not representative for the global market – a variance of 50% for the most popular names and over 90% for major proposed new names show the volatility of these numbers. It is highly doubtful the numbers presented in any such study should be the decisive factor for proceeding with or delaying the introduction of new gTLD name strings to the root – it remains a political decision.

Please feel free to contact us for any further information on the limited study concluded here in Germany over the last week:

eco Verband der deutschen Internetwirtschaft
Names and Numbers Forum
Klaus Landefeld, Member of the Board
Thomas Rickert, Director Names & Numbers

Lichtstr. 43h
50825 Köln
Germany