Study of the Prevalence of DNS Queries for CORP, HOME, and MAIL

ICANN Office of the Chief Technology Officer

Roy Arends

OCTO-007
14 April 2020

TABLE OF CONTENTS

This document is part of the OCTO document series. Please see https://www.icann.org/resources/pages/octo-publications-2019-05-24-en for a list of documents in the series. If you have questions or suggestions on any of these documents, please send them to octo@icann.org.

The data and analysis in this document were originally published internally to ICANN on 22 June 2017, and the data in this document is appropriate for that date. This document is being published in 2020 to enable wider community awareness of the analysis from that time.

1 Introduction

This work studies the prevalence of Domain Name System (DNS) queries for domains ending in the top-level labels CORP, HOME, and MAIL in traffic to the root servers over an extended period of time. The frequency of queries and the number of unique source addresses are measured and compared to queries for other non-existent domains.

The motivation for this work is to compare previous work in this area with more recent data, and also the ability to study trends over a longer period of time. The earlier work, *Name Collision in the DNS*[1] (commonly called "the Interisle report") is compared to this new analysis to determine if there are any significant changes in query patterns.

The Interisle report contains the following conclusion:

> "*For a broad range of potential policy decisions, a cluster of proposed TLDs at either end of the delegation risk spectrum are likely to be recognizable as "high risk" and "low risk." At the high end, the cluster includes the proposed TLDs that occur with at least order-of-magnitude greater frequency than any others (corp and home) and those that occur most frequently in internal X.509 public-key certificates (mail and exchange in addition to corp)."*

This study looks at queries at the time of writing (June 2017) and compares the collected statistics to the data collected in the Interisle report.

2 Methodology

To understand if there have been any changes to the volume of queries and volume of unique sources for CORP, HOME, and MAIL, the current research uses traffic data from two root server operators, B-root and L-root. The deployment of the two root server systems is notably different. At the time of analysis, L-root consists of 142 "anycast" instances, while B-root consists of one unicast system. Because server selection algorithms in resolvers determine how and when each of the 13 root server instances is queried, using two distinct sources will give a broader view of incoming queries to the root server system.

This study analyzes a complete data set capturing 19 months (from 1 September 2015 to 31 March 2017) of traffic sent to B-root. Additionally, there is a complete data set capturing over 9 months of traffic (from 1 September 2016 to 31 May 2017) sent to L-root. The analysis uses an interval of eight days, i.e., sampling data every eighth day, which is defined as UTC midnight to UTC midnight. This technique ensures that every day of the week is evenly sampled over the 18-month period of the collected B-root traffic. Lastly, the L-root data set is used as a control set: to confirm that the B-root traffic is representative, the relative ranking of domains in observed queries should be the same regardless of the number of unique sources and volume.

[1] See https://www.icann.org/en/system/files/files/name-collision-02aug13-en.pdf

Only queries from request sources that had the Recursion Desired (RD) bit in the DNS header cleared were considered. Some malware and some diagnostic tools (such as “dig”) set the RD bit by default. Only requests over User Datagram Protocol (UDP) were considered. Most TCP traffic observed was either zone transfer requests for the root zone or generated by clients who had tried UDP first.

2.1 The 2013 Interisle eport

Table 1 below is copied from the Interisle report. It shows the number (in thousands) of distinct IP address prefixes used to access each of the most queried proposed TLD strings for 2013, including the 2012 rank for comparison and the query count for each of the domains. These numbers are derived from DNS-OARC’s *Day In The Life* (DITL) root-server traffic data data in 2012 and 2013, which includes 24 hours of traffic to most of the root servers.[2] The focus of this study is the CORP, HOME, and MAIL domains, which are highlighted in red (rows 1, 2, and 22).

2013 rank	2012 rank	String	Count (thousands) 2013	Prefix Count (thousands) 2013
1	1	home	952,944	302
2	2	corp	144,507	185
3	21	ice	19,789	48
4	4	global	12,352	308
5	29	med	10,801	80
6	3	site	10,716	50
7	5	ads	10,563	148
8	12	network	8,711	57
9	7	group	8,580	45
10	9	cisco	8,284	78
11	8	box	7,694	89
12	14	prod	7,004	82
13	6	iinet	5,427	70
14	10	hsbc	5,249	90
15	11	inc	5,208	38
16	18	win	5,199	41
17	13	dev	5,058	104
18	15	office	4,006	88
19	20	business	3,279	59
20	16	host	3,127	98
21	31	star	2,435	88
22	25	mail	2,383	526
23	19	ltd	1,990	40
24	23	google	1,859	926
25	169	sap	1,735	41
26	17	app	1,720	112
27	27	world	1,650	24
28	30	mnet	1,568	37

[2] See https://www.dns-oarc.net/oarc/data/ditl

29	26	smart	1,331	38
30	33	web	1,126	191
31	32	orange	1,072	220
32	24	red	1,043	232
33	43	msd	956	11
34	37	school	872	28
35	39	bank	780	38

Table 1, a copy of “Table 6” from the Interisle report

Since 2013, most of the TLD labels in this table have been delegated. Due to the nature of caching in resolvers, delegated top-level domains are cached differently than non-existent top-level domains.[3] It is therefore not possible to compare the ranking of non-existent labels from 2012 and 2013 with the ranking of delegated labels in the current data set.

The Interisle report contains another table (table 3 in the original report) that shows the most frequently occurring top-level domains in the DITL 2013 data collection for all categories except “invalid”.[4] That table is shown below as Table 2. For the purpose of comparing data, Table 2 has been pruned: the domains that have been delegated since the publication of the Interisle report have been removed to show the ranking of the top 10 non-existent top-level domains. Note that the MAIL top-level domain is not present in the original table.

Interisle (2013)	**This report (2017)**	**TLD**	**Count (Interisle 2013) (thousands)**
1	1	local	2,501,349
2	3	home	1,018,998
3	6	localdomain	596,069
4	4	internal	508,937
5	22	localhost	414,286
6	7	belkin	388,979
7	5	lan	362,914
8	10	domain	275,608
9	8	corp	153,012
10	15	router	140,124

Table 2, pruned “Table 3” from the Interisle report, supplemented with the 2016 ranking.

This data shows that the ranking for HOME and CORP has not significantly changed between 2013 and 2017.

[3] Resolvers may cache records for a period of time set by the records’ “time to live” (TTL) value. Delegation point name server records in the root zone have a value of 2 days. Resolvers generally cache the non-existence of a record for a much shorter amount of time (e.g., 15 minutes or 1 hour, depending on the implementation). It is not possible to simply compare volumes of queries for existing names to volumes for non-existing names because of this difference in caching duration.

[4] The Interisle report considers top-level domains to be invalid if they do not comply with the rules specified in the Applicant Guide Book (https://newgtlds.icann.org/en/applicants/agb). For instance, top-level domains must be at least three characters long, and must only consist of alphabetical characters.

The rest of this paper will provide a deeper analysis. The analysis looks at the top 35 second-level domains for each of HOME, CORP, and MAIL, and then breaks them down in unique queries and unique requestors.

2.2 HOME, inverse order by volume (2017)

Rank	Requested String	Volume Observed	Average Daily Sources
1	hitronhub.home	355,980,000	8,809
2	com.home	330,355,963	23,396
3	net.home	90,881,094	14,512
4	wi-fiwalker.home	52,073,739	853
5	ru.home	25,535,657	4,492
6	cn.home	18,714,104	5,715
7	org.home	18,297,315	8,759
8	fios-router.home	17,810,784	8,146
9	_udp.home	15,275,108	12,317
10	wpad.home	13,486,870	18,682
11	isatap.home	10,970,354	24,802
12	3.home	10,802,005	2,169
13	_tcp.home	9,455,814	9,439
14	arpa.home	5,949,743	3,328
15	flybox.home	5,594,943	330
16	me.home	4,951,229	4,465
17	tv.home	4,699,858	4,520
18	info.home	4,410,235	3,631
19	kz.home	3,879,301	1,429
20	pl.home	3,661,169	2,434
21	biz.home	3,600,046	2,981
22	vn.home	3,535,300	992
23	in.home	3,166,481	2,309
24	cc.home	2,786,470	2,855
25	co.home	2,709,325	3,973
26	workgroup.home	2,586,929	6,029
27	de.home	2,380,916	2,949
28	.home	2,178,538	5,737
29	io.home	2,176,382	4,166
30	ir.home	2,102,179	1,328
31	home.home	2,054,884	4,411
32	it.home	2,022,778	2,610
33	jp.home	1,903,484	2,923
34	eu.home	1,856,928	2,327
35	m.home	1,827,172	2,339

Table 3, second-level domains under HOME, by volume.

Table 3 shows the top 35 queries for domains under the HOME top-level domain, ordered by volume, observed over a period of 585 days, sampled every eighth day. The domains are aggregated by second-level domain. For instance, “example.com.home” is aggregated under "com.home". The red entries in the Requested String column identifies a string where the second-level domain exists as a top-level domain. The large amount of top-level domain labels as second-level is typical for networks that have the string "HOME" configured as a search domain.

2.3 HOME, order by the unique number of sources (2017)

Rank	Requested String	Volume Observed	Average Daily Sources
11	isatap.home	10,970,354	24,802
2	com.home	330,355,963	23,396
10	wpad.home	13,486,870	18,682
3	net.home	90,881,094	14,512
9	_udp.home	15,275,108	12,317
13	_tcp.home	9,455,814	9,439
1	hitronhub.home	355,980,000	8,809
7	org.home	18,297,315	8,759
8	fios-router.home	17,810,784	8,146
26	workgroup.home	2,586,929	6,029
28	.home	2,178,538	5,737
6	cn.home	18,714,104	5,715
17	tv.home	4,699,858	4,520
5	ru.home	25,535,657	4,492
60	retracker.home	917,512	4,475
16	me.home	4,951,229	4,465
31	home.home	2,054,884	4,411
29	io.home	2,176,382	4,166
25	co.home	2,709,325	3,973
18	info.home	4,410,235	3,631
14	arpa.home	5,949,743	3,328
21	biz.home	3,600,046	2,981
27	de.home	2,380,916	2,949
33	jp.home	1,903,484	2,923
24	cc.home	2,786,470	2,855
36	to.home	1,792,995	2,645
32	it.home	2,022,778	2,610
50	gov.home	1,220,428	2,539
38	us.home	1,742,866	2,534
49	local.home	1,232,438	2,467
40	uk.home	1,531,383	2,465
20	pl.home	3,661,169	2,434
54	mobi.home	979,745	2,382
87	http.home	449,135	2,349
35	m.home	1,827,172	2,339

Table 4, second-level domains under HOME, by number of unique source addresses.

Table 4 shows the top 35 queries, ordered by Average Daily Sources with the ranking from Table 3. The red entries in the Requested String column identifies a string that lies outside of the previous volume table ranking. Again, quite a few top-level domains as second-level domains are observed. The first and second entries see about the same number of unique sources, but at a substantially different volume.

2.4 CORP, inverse order by volume (2017)

Rank	Requested String	Volume Observed	Average Daily Sources
1	bank.corp	42,059,123	4,850
2	sap.corp	11,664,894	7,835
3	ecolab.corp	11,517,301	6,907
4	compassgroup.corp	10,631,376	3,828
5	zurich.corp	10,108,509	3,758
6	cam.corp	9,860,351	4,748
7	bvcorp.corp	8,679,480	5,314
8	guardian.corp	8,044,236	1,054
9	parker.corp	6,303,956	5,156
10	sungard.corp	6,106,491	2,954
11	root.corp	5,940,672	3,394
12	teva.corp	5,029,533	3,771
13	davita.corp	4,979,549	2,280
14	airbus.corp	4,564,945	2,668
15	internal.corp	4,411,798	3,033
16	sanm.corp	4,362,007	1,932
17	quest.corp	3,699,212	2,420
18	global.corp	2,655,743	2,318
19	alico.corp	2,641,452	2,423
20	bmw.corp	2,546,432	3,513
21	stream.corp	2,381,950	1,828
22	sealedair.corp	2,320,861	3,339
23	hospira.corp	2,245,749	2,960
24	ad.corp	2,243,920	2,409
25	abacus.corp	2,238,154	1,562
26	mbci.corp	2,204,625	1,048
27	logistics.corp	2,076,638	2,901
28	delta.corp	1,974,371	2,538
29	directenergy.corp	1,832,363	1,654
30	sdl.corp	1,768,486	1,624
31	bi.corp	1,696,800	1,389
32	abg.corp	1,601,522	2,073
33	eurocopter.corp	1,589,311	1,260
34	hrc.corp	1,553,590	2,000
35	us.corp	1,409,403	1,639

Table 5, second-level domains under CORP, by volume.

Table 5 shows the top 35 queries for domains under the CORP top-level domain, ordered by volume, observed over a period of 585 days, sampled every eighth day. The domains are aggregated by second-level domain. For instance, “example.com.corp” is aggregated under “com.corp”. A large part of this top 35 list consists of recognizable global brands. There is no significant presence of top-level domains as second-level domains. This suggests that the CORP domain is configured mainly in search domains in corporations around the world and that this domain might be in active local use at these corporations.

2.5 CORP, order by the unique number of sources (2017)

Rank	Requested String	Volume Observed	Average Daily Sources
2	sap.corp	11,664,894	7,835
3	ecolab.corp	11,517,301	6,907
7	bvcorp.corp	8,679,480	5,314
9	parker.corp	6,303,956	5,156
1	bank.corp	42,059,123	4,850
6	cam.corp	9,860,351	4,748
4	compassgroup.corp	10,631,376	3,828
12	teva.corp	5,029,533	3,771
5	zurich.corp	10,108,509	3,758
20	bmw.corp	2,546,432	3,513
11	root.corp	5,940,672	3,394
22	sealedair.corp	2,320,861	3,339
15	internal.corp	4,411,798	3,033
23	hospira.corp	2,245,749	2,960
10	sungard.corp	6,106,491	2,954
27	logistics.corp	2,076,638	2,901
14	airbus.corp	4,564,945	2,668
28	delta.corp	1,974,371	2,538
19	alico.corp	2,641,452	2,423
17	quest.corp	3,699,212	2,420
24	ad.corp	2,243,920	2,409
18	global.corp	2,655,743	2,318
13	davita.corp	4,979,549	2,280
63	.corp	527,824	2,154
32	abg.corp	1,601,522	2,073
38	hsi.corp	1,352,153	2,022
34	hrc.corp	1,553,590	2,000
69	abbott.corp	351,555	1,940
16	sanm.corp	4,362,007	1,932
36	hologic.corp	1,368,183	1,910
21	stream.corp	2,381,950	1,828
45	brkr.corp	974,774	1,758

29	directenergy.corp	1,832,363	1,654
35	us.corp	1,409,403	1,639
30	sdl.corp	1,768,486	1,624

Table 6, second-level domains under CORP, by number of unique source addresses.

Table 6 shows the top 35 queries for CORP ordered by average daily sources and includes the ranking from Table 5. The red entries in the column of requested strings identify a string that lies outside the ranking in Table 5. Again, a few recognizable global brand names are observed. The average number of daily sources is lower than that of HOME sources. CORP domains are likely used in corporate environments, related to the second-level domains. HOME domains are likely used in home environments, which are unrelated to the second-level domain.

2.6 MAIL, reverse order by volume (2017)

Rank	Requested string	Volume Observed	Average Daily Sources
1	.mail	8,496,910	10,941
2	system.mail	361,694	2,265
3	win.mail	357,709	1,417
4	alico.mail	350,051	796
5	al.mail	187,074	367
6	g.mail	173,779	1,054
7	yahoo.mail	145,612	1,094
8	com.mail	105,209	334
9	hot.mail	84,580	450
10	mail.mail	80,488	451
11	google.mail	55,151	263
12	company.mail	54,168	432
13	gmail.mail	53,229	238
14	navy.mail	50,687	193
15	army.mail	44,085	192
16	_tcp.mail	42,754	280
17	_sites.mail	41,395	261
18	infra.mail	38,370	61
19	net.mail	34,954	160
20	ct.mail	34,833	38
21	af.mail	34,639	133
22	aol.mail	34,228	211
23	www.mail	34,068	325
24	hotmail.mail	31,627	152
25	winus.mail	29,264	182
26	sw.mail	27,441	10
27	e.mail	26,351	218
28	receive.mail	24,005	124
29	maillocal.mail	20,768	63
30	smtp.mail	20,234	149
31	cra.mail	19,890	177
32	embarq.mail	18,492	81

33	rocket.mail	17,613	93
34	tp.mail	16,787	7
35	yandex.mail	16,579	84

Table 7, second-level domains under MAIL, by volume.

Table 7 shows the top 35 queries for domains under the MAIL top-level domain. There are a significant number of second-level domains that are similar to popular web-based email hosting services, such as g.mail, yahoo.mail, gmail.mail and hot.mail. This table shows, both in volume and unique sources, less traffic than HOME and CORP. The MAIL top-level domain does not appear in the current top 10 list of non-existent top-level domains. In the 2012 and 2013 DITL traffic data, the MAIL top-level domain appeared at rank 22, lower than other domains that have been delegated since. Additionally, this traffic may show mistyped domains for the HOTMAIL, GMAIL, and EMAIL top-level domains. The Interisle report mentions that MAIL was the highest non-delegated top-level domain in traffic to the resolver. This study does not include resolver traffic.

2.7 MAIL, order by the unique number of sources (2017)

Rank	Requested string	Volume Observed	Average Daily Sources
1	.mail	8,496,910	10,941
2	system.mail	361,694	2,265
3	win.mail	357,709	1,417
7	yahoo.mail	145,612	1,094
6	g.mail	173,779	1,054
4	alico.mail	350,051	796
10	mail.mail	80,488	451
9	hot.mail	84,580	450
12	company.mail	54,168	432
5	al.mail	187,074	367
8	com.mail	105,209	334
23	www.mail	34,068	325
16	_tcp.mail	42,754	280
11	google.mail	55,151	263
17	_sites.mail	41,395	261
13	gmail.mail	53,229	238
27	e.mail	26,351	218
22	aol.mail	34,228	211
14	navy.mail	50,687	193
15	army.mail	44,085	192
25	winus.mail	29,264	182
31	cra.mail	19,890	177
19	net.mail	34,954	160
24	hotmail.mail	31,627	152
30	smtp.mail	20,234	149
21	af.mail	34,639	133
28	receive.mail	24,005	124

40	qq.mail	13,057	105
39	imap.mail	13,896	99
33	rocket.mail	17,613	93
38	live.mail	14,131	87
36	mil.mail	16,394	85
35	yandex.mail	16,579	84
32	embarq.mail	18,492	81
41	_msdcs.mail	11,422	77

Table 8, second-level domains under MAIL, by number of unique source addresses.

Table 8 shows the top 35 queries, ordered by average daily sources, and includes the ranking from Table 7. The red entries in the column of requested strings identifies a string that lies outside the ranking in Table 7.

2.8 Validating observations using a control

In this study, traffic from L-root is used as a control. Seven individual days have been compared between B-root and L-root for both CORP and HOME domains. Due to the larger number of instances, L-root has a higher volume and more unique query sources than the B-root. However, the average ranking of requested strings remains mostly the same. To highlight this finding, the next table shows the values and ranking for both B-root and L-root for data from the exact same UTC day (5 October 2016).

Rank for L	Rank for B	Requested String	Volume B	Volume L	Volume Ratio	Uniq Src B	Uniq Src L	Uniq Src Ratio
1	1	bank.corp	456,170	1,530,975	3.36	5,445	6,389	1.17
2	3	cam.corp	170,376	938,210	5.51	6,784	9,616	1.42
3	2	sap.corp	210,305	697,298	3.32	8,607	15,109	1.76
4	4	ecolab.corp	163,124	635,469	3.90	7,645	11,771	1.54
5	7	bvcorp.corp	107,831	550,608	5.11	5,627	9,220	1.64
6	11	airbus.corp	65,554	449,454	6.86	3,167	5,507	1.74
7	9	parker.corp	67,329	366,015	5.44	6,852	12,009	1.75
8	8	zurich.corp	94,504	328,046	3.47	4,060	5,723	1.41
9	5	compassgroup .corp	134,113	314,798	2.35	4,463	4,706	1.05
10	12	davita.corp	56,372	279,524	4.96	2,556	2,840	1.11
11	10	sungard.corp	66,861	245,994	3.68	3,019	4,344	1.44
12	14	teva.corp	52,026	243,710	4.68	3,963	6,641	1.68
13	29	global.corp	22,762	223,427	9.82	2,476	4,712	1.90
14	16	internal.corp	48,255	219,543	4.55	3,081	5,435	1.76
15	13	root.corp	52,474	206,347	3.93	3,519	5,490	1.56
16	20	sanm.corp	38,885	194,584	5.00	2,027	3,075	1.52
17	25	ad.corp	24,544	180,811	7.37	2,580	4,779	1.85
18	48	rackspace.corp	12,270	160,255	13.06	1,269	1,837	1.45
19	15	alico.corp	50,035	157,026	3.14	3,256	5,371	1.65
20	19	quest.corp	41,286	155,867	3.78	2,737	4,806	1.76

21	30	eurocopter.corp	21,488	148,570	6.91	1,349	2,409	1.79
22	24	logistics.corp	26,283	137,577	5.23	3,191	5,430	1.70
23	21	bmw.corp	31,681	134,488	4.25	3,743	6,839	1.83
24	18	webtrends.corp	42,876	132,217	3.08	294	508	1.73
25	54	zh.corp	10,319	128,217	12.43	435	684	1.57
26	6	guardian.corp	127,366	127,488	1.00	943	1,811	1.92
27	23	sealedair.corp	27,088	127,232	4.70	3,371	6,223	1.85
28	38	bi.corp	16,144	103,883	6.43	1,510	2,014	1.33
29	28	directenergy.corp	23,458	100,304	4.28	2,005	2,393	1.19
30	35	sdl.corp	18,823	99,320	5.28	1,589	2,849	1.79

Table 9, second-level domains under CORP, observed in L-root, ranked by B-root.

Table 9 is ordered by the volume observed for L-root (the fifth column, “Volume L”). The second column (“Ranking for L”) shows the ranking that the observed domains would have if the table were ordered by the volume observed for B-root (the fourth column, “Volume B”). The average ratio (“Volume Ratio”) between the volume of B-root and L-root is 1:4, i.e., L-root receives four times the traffic that B-root receives. Additionally, L-root has about 60% more unique source addresses compared to B-root. Though L-root sees a higher number than B-root in both unique source addresses and volume, this disparity does not influence the overall ranking. Notably, in the top 30 of observed domains in the traffic for L-root, 26 domains appear in the top 30 of observed domains in the traffic for B-root. The remaining four domains have ranking 35, 38, 48, and 54 in the L-root traffic. This stability in ranking between B-root and L-root was also observed in the ranking for the HOME domain between B-root and L-root.

3 Conclusion

CORP and HOME remain among the most requested top-level domains. The number of average daily sources shows how wide these top-level domains are in use. Ranking of CORP and HOME top-level domains remains very high, both in volume and number of unique source addresses. There has been no significant change in the ranking of CORP and HOME in the observed traffic between the 2012 and 2013 DITL observations and this longitudinal study.

Since the ranking of the MAIL top-level domain in root server traffic has not changed, it should not be assumed that the ranking of the MAIL top-level domain in resolver traffic has changed. Further study of resolver data is needed to determine if the MAIL top-level domain ranking in resolver traffic has changed.