Personal comment on the ICANN staff proposal to mitigate name collisions

Real-world observations based on Corp.com query traffic

Mikey O'Connor

I decided to pretend to be a new-gTLD applicant and research the query traffic coming to my corp.com domain. The goal of this research is to simulate what an applicant (turning into a registry) might encounter when they attempt to follow the ICANN staff proposal to mitigate name collisions in a namespace like corp.com.

The short summary of what I found

The data. A few snippets are contained in the body of this memo; more detailed (but still summary) data is attached in the Appendix.

The conclusion. I am delighted not to be an applicant facing the job that ICANN is proposing – as it's currently proposed, I don't think it's possible to do it.

The way forward. I am very interested in collaborating with others (researchers, working groups, vendors, network operators, etc.) to better understand the risks that are implied by the traffic to corp.com and figure out how to mitigate those risks.

Here's the long version...

Approach to the study

I think I typify the average new-gTLD applicant in that my DNS and log-analysis skills really aren't up to this job. So I worked with Interisle Consulting Group to do the crash-speed project of setting up an authoritative DNS server, extracting queries, and analyzing query log data. My sincere thanks to Lyman Chapin, Jim Reid and Colin Strutt for the work they did (and the good humor with which they approached it) over this past weekend.

Much more work needs to be done to understand the risks that would occur if I, the imaginary registry, were to start delegating names under corp.com – and, while I'm not going to pursue that kind of research on my own, I would be a willing partner in such an effort if others (SSAC? Microsoft? DNS-OARC? Verisign?) were interested.

I'm not going to go into huge detail on how Interisle did all this work. I can defend this summary of our data; I just don't have time to write the paragraphs before the document deadline. In summary, this study is based on a very rough-cut review of

about 48 hours worth of qlog data from an authoritative DNS server, answering queries to corp.com.

The hope

My hope was that the query traffic would show me an easy problem to fix. I was looking for:

- Small amounts of traffic,
- Queries for easy to identify services,
- Queries from a small number of easy to identify organizations and IP addresses.

None of these dreams came true. **Much** more work is required to truly understand the implications of the traffic that we're seeing.

Small amounts of traffic?

In a word, no.

Corp.com gets between 1.5 and 2 million queries a day (see Appendix A), which is a pretty substantial proportion of the 122 million queries that Interisle lists for the applied-for .corp in their report. This is especially true since the 122 million queries listed for .corp were recorded over a 48 hour period – that translates to roughly 60 million queries a day. Thus corp.com receives about 3-4% of the traffic of .corp.

This also puts corp.com around 20th on the list in the Interisle report, if we were to pretend that it's a TLD included in the Interisle study. So much for a "small amount of traffic."

Queries for easy to identify services?

The Interisle team broke the corp.com queries down into four "interesting clumps" in this rough-cut analysis. As a result, the numbers won't all add up to the same total as there are overlaps. Here are the clumps:

- 2 dots in the name (for example, foo.corp.com)
- Underscore in the name (usually, but not always, some sort of service name)
- A name containing the string "isatap." or "wpad." somewhere in the name
- The rest (i.e., other than 2 dots, no underscore, not including isatap. or wpad.)

A "good" answer would have found that the vast majority of the names are the simple "two dots in the name" queries. Unfortunately, only about half the queries fall in the "two dot" category and quite a few look to be queries for services of one sort or another. Here's the summary list:

- Two dots – 1,668,354 queries
- Underscore – 701,710 queries
- Isatap or Wpad – 129,536 queries
- The rest – 643,548 queries

Queries from a small number of easy to identify organizations and IP addresses?

A "good" answer would have found that the queries were concentrated in just a few IP addresses, so I would only have to contact a few organizations. Unfortunately this is also not the case. First, here's a count of the number of IP addresses:

- Two dots – 20,178 addresses
- Underscore – 11,968 addresses
- Isatap or Wpad – 8,804 addresses
- The rest – 18,300 addresses

And here's how little they concentrate:

- Two dots – 28 account for 10%; 600 account for 50% of IP addresses
- Underscore – 33 account for 10%; 674 account for 50% of IP addresses
- Isatap or Wpad – 57 account for 10%; 891 account for 50% of IP addresses
- The rest – 11 account for 10%; 440 account for 50% of IP addresses

So if I were an applicant turning into a registry, I would have to track down a lot of separate IP addresses just to cover 50% of the traffic. Oh, and did I mention that these queries are coming from something over 130 countries? Here's the list of the top three per Country Code (CC):

- Two dots – 136 CCs, US (35.6%), CN (6.5%), DE (6.0%)
- Underscore – 94 CCs, US (44.2%), CN (5.3%), TW (4.3%)
- Isatap or Wpad – 92 CCs, US (47.5%), TW (9.1%), CN (4.7%)
- The rest – 124 CCs, US (39.1%), BE (17.8%), CN (4.8%)

Feasibility of the ICANN mitigation approach

Again pretending to be an applicant, notifying the sources of the colliding traffic looks like a lot of very tricky work to do – especially because there are several factors that are likely to make the research and notification work more difficult:

- The RIRs are suggesting that this is an inappropriate use of their Whois data – so where exactly would I go to find out who's connected to the IP addresses I'm seeing?
- Much of this traffic is coming from intermediate resolvers (ISPs, DNS providers and application-providers such as Google) – those folks might not greet me with enthusiasm if I were to contact them and ask them to identify the ultimate source of the queries. If they could do it at all.
- There may be several DNS hops between what I see and the actual originator of the traffic – so all these counts may be understating the work by quite a bit
- The organizations I would be asking to research the source of the queries might want to be compensated for the time and effort it takes for them to do the research – a completely unbudgeted project for both of us.
- Getting all this research and mitigation organized in the proposed 30-day window would be challenging, to say the least.

What's next?

There are several questions on my mind as I write this comment.

- Is there a way to better understand the nature of the risk and the source of the traffic, especially given some of those constraints?
- Is there a way that application and hardware providers (or their VARs) could help with the assessment and mitigation?
- Is there a way that others might help with the assessment and mitigation of other traffic relating to service discovery?

Conclusions

As I end my brief time pretending to be a new gTLD applicant, I come away doubting the feasibility of using raw query-traffic data to answer the questions "what is the nature of the risk posed by this traffic?" and "who can I contact to alert them of the risk?" This data is much "shallower" than I thought it would be. Sure, there's a lot of it, but it suffers from all the shortcomings I've listed above. It doesn't precisely identify what is likely to break, nor does it tell me whom to inform. It's clear to me that the analysts of the data will have to make informed guesses about what the data is for and where it is coming from. That will not be a simple or speedy process.

Presuming away those troubles for a moment, I'm overwhelmed by the amount of tedious notification and mitigation work that would be required. Let's say that

there are perfect ID's of the problems, and the people who have them, what then? Does the applicant/registry/registry-service-provider send them unsolicited email? Call them on the phone? Launch a big PR campaign warning people that their systems may break? Put all the colliding strings on a reserved-name list?

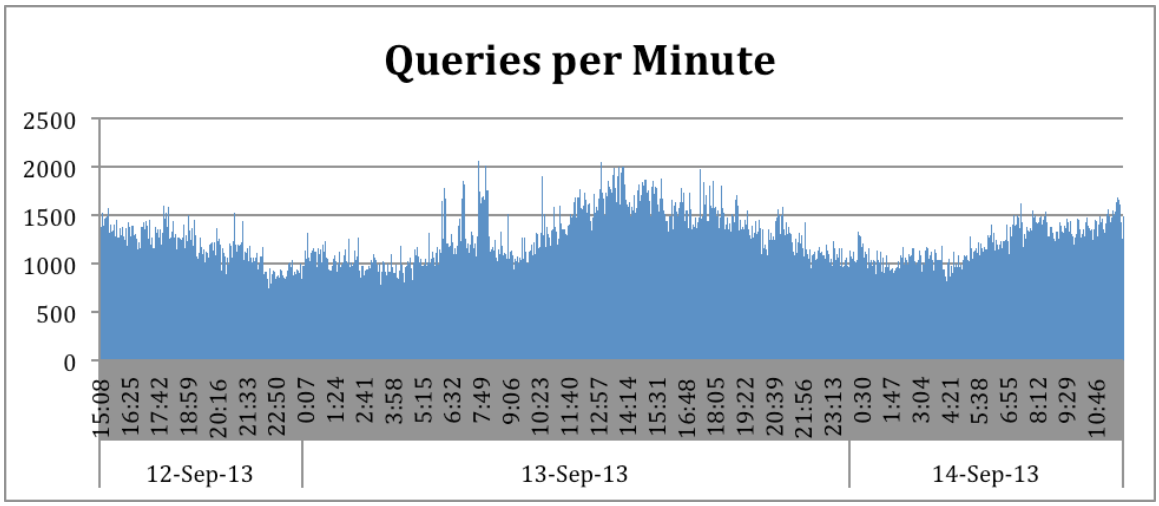
And what happens once I've reached those people? How do I begin to help the tens, maybe hundreds, of thousands of network administrators with their mitigation efforts? What happens if the advice I give them is bad and things break as a result of my advice? How do I manage all those relationships, none of which are revenue bearing?

What are the alternatives to doing all that? I now have a lot of data describing this problem – it doesn't seem appropriate to just shift the work to the people whose networks will break and say, "tra-la-la, you fix it."

Fortunately, I can now pull off my imaginary-registry hat and put my retired-guy hat back on. There's a lot of stuff in this name-collision conversation that goes on the "too hard" pile for me.

I'm delighted not to be an applicant facing the job that ICANN is proposing for me – as proposed, I don't think it can be done.

Appendix A – Data summaries



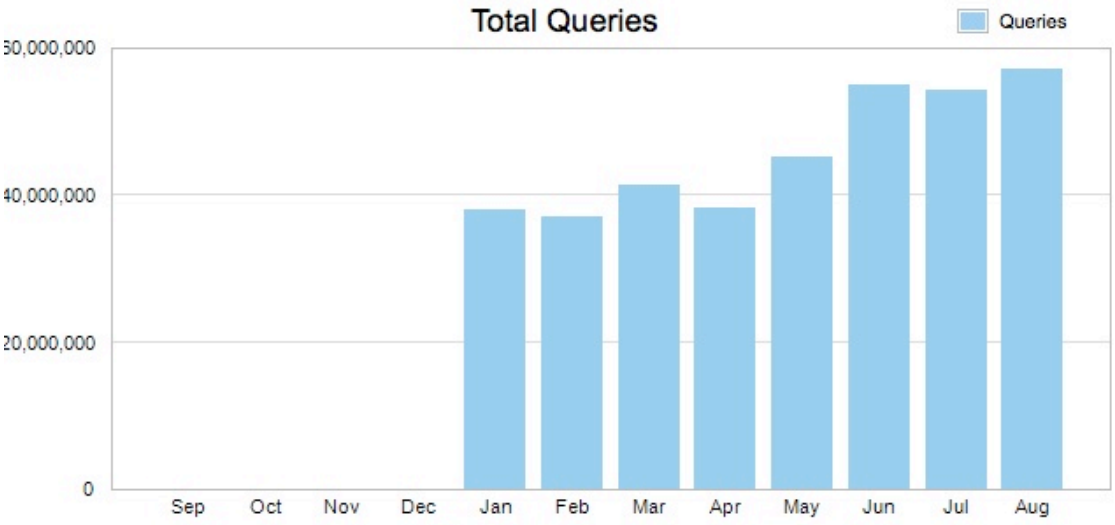
This is per-minute query data in the **sample**. Note: we'd turned TTL down from 1 hour to 1 week, so this is a sample of a downward-trending line as the TTL propogated.

The 1.5 to 2 million queries a day estimate is based on this earlier report from Godaddy's Advanced DNS, the authoritative DNS server before switching to our own authoritative server. The June increase seems to be due to a change in Godaddy metrics, as a similar increase appears across several other high-trafic names that were available for comparison.

Today	30-Day Average	This Month	12-Month Average
1,590,000 Queries	1,849,673/day	10320000 Queries	10320000/mo

Export

9/1/2012 - 8/31/2013



Date	Queries
September 2012	0
October 2012	0
November 2012	0
December 2012	0
January 2013	37908068
February 2013	37109032
March 2013	41328318
April 2013	38199588
May 2013	45049805
June 2013	55018435
July 2013	54290461
August 2013	56996242
Total	**365899949**

Note: All of the following tables show the "top twenty" in each case.

Queries by name

All queries		
QNAME	Queries	Cum %
corp.co.corp.com	140,893	4.5%
_ldap._tcp.dc._msdcs.corp.com	66,128	6.6%
ns1.corp.com	60,405	8.5%
ns2.corp.com	60,273	10.4%
_ldap._tcp.ef84b3da-3710-46ed-9936-a7a7b39d8be3.domains._msdcs.corp.com	55,261	12.1%
wpad.corp.com	51,317	13.7%
_ldap._tcp.corp.com	33,278	14.8%
_kerberos._tcp.dc._msdcs.corp.com	27,973	15.7%
invsfoxepo01.corp.com	24,254	16.5%
_ldap._tcp.gc._msdcs.corp.com	21,509	17.1%
_ldap._tcp.pdc._msdcs.corp.com	20,871	17.8%
sms_slp.corp.com	15,438	18.3%
isatap.corp.com	14,350	18.7%
usfxbinvscm01pv.corp.com	13,978	19.2%
invsfoxscm02.corp.com	12,866	19.6%
invshouxchmbx04.corp.com	11,731	20.0%
invsfoxxchmbx02.corp.com	11,061	20.3%
invshouxchpub01.corp.com	10,921	20.7%
_ldap._tcp.foxboro._sites.dc._msdcs.corp.com	10,819	21.0%
invsfoxxchpub01.corp.com	10,605	21.3%

Two dots		
QNAME	Queries	Cum %
ns1.corp.com	60,405	3.6%
ns2.corp.com	60,273	7.2%
invsfoxepo01.corp.com	24,254	8.7%
sms_slp.corp.com	15,438	9.6%
usfxbinvscm01pv.corp.com	13,978	10.5%
invsfoxscm02.corp.com	12,866	11.2%
invshouxchmbx04.corp.com	11,731	11.9%
invsfoxxchmbx02.corp.com	11,061	12.6%
invshouxchpub01.corp.com	10,921	13.2%

invsfoxxchpub01.corp.com	10,605	13.9%
invsfoxxchmbx01.corp.com	10,421	14.5%
invsfoxepo02v.corp.com	9,871	15.1%
invshouxchmbx03.corp.com	9,554	15.7%
invslkfscm01.corp.com	9,334	16.2%
dedussrvfil001.corp.com	9,198	16.8%
invssclscm01.corp.com	9,052	17.3%
ctlman0012.corp.com	9,020	17.9%
rumosiomfil01pp.corp.com	8,944	18.4%
inchnsrvapp008.corp.com	8,904	18.9%
ctlbelfil01.corp.com	8,891	19.5%

Underscore		
QNAME	Queries	Cum %
_ldap._tcp.dc._msdcs.corp.com	66,128	9.4%
_ldap._tcp.ef84b3da-3710-46ed-9936-a7a7b39d8be3.domains._msdcs.corp.com	55,261	17.3%
_ldap._tcp.corp.com	33,278	22.0%
_kerberos._tcp.dc._msdcs.corp.com	27,973	26.0%
_ldap._tcp.gc._msdcs.corp.com	21,509	29.1%
_ldap._tcp.pdc._msdcs.corp.com	20,871	32.1%
sms_slp.corp.com	15,438	34.3%
_ldap._tcp.foxboro._sites.dc._msdcs.corp.com	10,819	35.8%
_ldap._tcp.foxboro._sites.corp.com	7,715	36.9%
_ldap._tcp.london._sites.dc._msdcs.corp.com	5,560	37.7%
_ldap._tcp.dc._msdcs.benq.corp.com	5,000	38.4%
_ldap._tcp.b274bfe3-0bb4-40c5-a4f7-069e5b43d57c.domains._msdcs.corp.com	4,469	39.1%
_mssms_mp_inv._tcp.corp.com	4,409	39.7%
_ldap._tcp.default-first-site-name._sites.dc._msdcs.corp.com	3,968	40.2%
nlb_inv.corp.com	3,850	40.8%
mp_inv.corp.com	3,786	41.3%
_ldap._tcp.dc._msdcs.win.corp.com	3,747	41.9%
_ldap._tcp.london._sites.corp.com	3,669	42.4%
_ldap._tcp.lakeforest._sites.dc._msdcs.corp.com	3,574	42.9%
_kerberos._tcp.foxboro._sites.dc._msdcs.corp.com	3,447	43.4%

Isatap or Wpad		
QNAME	Queries	Cum %
wpad.corp.com	51,317	39.6%

isatap.corp.com	14,350	50.7%
wpad.benq.corp.com	4,383	54.1%
isatap.benq.corp.com	3,098	56.5%
wpad.win.corp.com	2,739	58.6%
wpad.accent.corp.com	2,476	60.5%
wpad.wm.corp.com	1,941	62.0%
wpad.bqc.corp.com	1,630	63.3%
isatap.accent.corp.com	1,622	64.5%
wpad.tsi.corp.com	1,540	65.7%
wpad.bqa.corp.com	1,501	66.9%
isatap.win.corp.com	1,420	67.9%
wpad.trx.corp.com	1,391	69.0%
wpad.bluerhino.corp.com	1,378	70.1%
wpad.dmz.trx.corp.com	1,274	71.1%
wpad.c00.corp.com	1,219	72.0%
wpad.mwp.corp.com	1,182	72.9%
wpad.usrenalcare.corp.com	1,098	73.8%
wpad.bluene.corp.com	1,063	74.6%
wpad.ce.corp.com	1,041	75.4%

The rest		
QNAME	**Queries**	**Cum %**
corp.co.corp.com	140,893	21.9%
sv1365.tsi.corp.com	6,400	22.9%
pdawin01.sprint.corp.com	4,528	23.6%
xsc03923.win.corp.com	4,067	24.2%
xsc03924.win.corp.com	3,949	24.8%
xp076596.win.corp.com	3,216	25.3%
hpserver.benq.corp.com	2,452	25.7%
symantec.win.corp.com	2,101	26.0%
xp064479.win.corp.com	2,046	26.4%
prmwntjo.ce.corp.com	2,004	26.7%
dpdcb.melinda.local.win.corp.com	2,002	27.0%
ecbes.alv.corp.com	2,000	27.3%
epic.ce.corp.com	1,956	27.6%
xs000800.win.corp.com	1,883	27.9%
pvs-av01.usrenalcare.corp.com	1,828	28.2%
wsus.benq.corp.com	1,669	28.4%
nb-bqca-tram-w8.bqa.corp.com	1,662	28.7%
mbx.ce.corp.com	1,660	29.0%
xp077531.win.corp.com	1,589	29.2%
xp074301.win.corp.com	1,586	29.4%

Queries by IP address

All queries			
IP	CC	Queries	%
195.34.133.21	AT	14,880	0.5%
177.39.193.66	BR	13,154	0.9%
64.233.162.80	US	12,430	1.3%
64.233.162.81	US	12,283	1.7%
64.233.162.82	US	12,270	2.1%
202.188.1.10	MY	10,876	2.4%
212.77.192.41	QA	9,701	2.7%
163.121.211.82	EG	9,443	3.0%
163.121.200.206	EG	9,334	3.3%
212.77.192.42	QA	9,095	3.6%
195.238.24.110	BE	8,318	3.9%
195.238.24.109	BE	8,314	4.1%
212.93.192.11	SA	8,113	4.4%
212.77.192.43	QA	8,026	4.6%
12.218.100.2	US	7,733	4.9%
213.224.146.19	BE	7,712	5.1%
213.224.146.52	BE	7,630	5.4%
213.224.146.20	BE	7,559	5.6%
213.224.146.51	BE	7,381	5.8%
195.238.25.110	BE	7,075	6.1%

Two dots			
IP	CC	Queries	%
195.34.133.21	AT	12,624	0.8%
64.233.162.81	US	9,062	1.3%
64.233.162.80	US	8,990	1.8%
64.233.162.82	US	8,792	2.4%
212.77.192.41	QA	8,162	2.9%
212.93.192.11	SA	8,089	3.3%
212.77.192.42	QA	7,655	3.8%
12.218.100.2	US	7,401	4.2%
202.188.1.10	MY	7,218	4.7%
212.77.192.43	QA	6,782	5.1%
200.33.148.210	MX	6,407	5.5%
201.144.127.241	MX	6,280	5.8%
208.76.26.4	US	5,901	6.2%
86.51.32.40	SA	5,099	6.5%
86.51.32.38	SA	4,856	6.8%

163.121.200.206	EG	4,851	7.1%
86.51.32.41	SA	4,822	7.4%
86.51.32.37	SA	4,818	7.7%
86.51.32.39	SA	4,803	7.9%
219.141.148.39	CN	4,612	8.2%

Underscore			
IP	**CC**	**Queries**	**%**
163.121.211.82	EG	4,970	0.7%
115.178.100.147	IN	4,528	1.4%
163.121.200.206	EG	4,394	2.0%
166.147.121.9	US	3,407	2.5%
202.188.1.10	MY	3,294	2.9%
163.121.213.180	EG	2,843	3.3%
163.121.128.90	EG	2,475	3.7%
81.92.223.36	PT	2,369	4.0%
195.34.133.21	AT	2,238	4.3%
163.121.128.94	EG	2,218	4.7%
65.55.125.40	US	2,089	5.0%
202.151.64.137	GU	2,067	5.3%
163.121.213.185	EG	2,031	5.5%
219.141.148.39	CN	1,815	5.8%
163.121.218.189	EG	1,768	6.1%
198.184.235.17	US	1,741	6.3%
121.253.7.31	KR	1,712	6.5%
64.202.160.225	US	1,693	6.8%
202.151.64.110	GU	1,689	7.0%
64.233.162.82	US	1,626	7.3%

Isatap or Wpad			
IP	**CC**	**Queries**	**%**
166.147.121.9	US	1,060	0.8%
202.151.64.110	GU	640	1.3%
202.151.64.137	GU	512	1.7%
65.38.111.37	US	319	2.0%
208.104.2.37	US	271	2.2%
123.176.37.37	IN	271	2.4%
166.147.73.1	US	269	2.6%
68.105.29.112	US	258	2.8%
15.219.145.213	US	254	3.0%
68.105.29.111	US	254	3.2%
65.55.125.41	US	252	3.4%

68.105.29.107	US	251	3.6%
68.105.29.110	US	251	3.8%
68.105.29.108	US	250	3.9%
68.105.29.109	US	241	4.1%
202.188.1.10	MY	234	4.3%
15.219.145.212	US	229	4.5%
65.117.207.137	US	228	4.7%
70.88.239.193	US	214	4.8%
68.105.29.144	US	209	5.0%

The rest			
IP	**CC**	**Queries**	**%**
177.39.193.66	BR	13,154	2.0%
195.238.24.109	BE	6,890	3.1%
195.238.24.110	BE	6,887	4.2%
195.238.25.110	BE	5,877	5.1%
195.238.25.108	BE	5,779	6.0%
195.238.25.109	BE	5,671	6.9%
98.159.4.4	US	4,919	7.6%
190.143.160.42	GT	4,107	8.3%
110.174.198.133	AU	3,873	8.9%
195.238.24.114	BE	3,528	9.4%
195.238.25.115	BE	3,385	10.0%
195.238.25.114	BE	3,366	10.5%
216.170.157.5	US	3,272	11.0%
195.238.25.116	BE	3,239	11.5%
195.238.24.115	BE	3,239	12.0%
195.130.131.10	BE	3,234	12.5%
195.238.24.116	BE	3,213	13.0%
195.238.24.108	BE	3,203	13.5%
213.224.146.19	BE	3,158	14.0%
213.224.146.52	BE	3,096	14.5%

Queries by 3rd-level name

3LD	Queries	IP subnets	IP addresses	%
_msdcs	322,566	1,096	9,697	10.2%
win	171,096	169	905	5.4%
co	140,950	2,595	12,467	4.5%
ns1	63,173	1,049	4,405	2.0%
ns2	63,045	1,040	4,383	2.0%
benq	53,370	316	2,074	1.7%
wpad	51,317	1,034	6,828	1.6%
_tcp	47,725	796	5,343	1.5%
ce	40,664	137	1,278	1.3%
tsi	40,355	131	1,023	1.3%
_sites	38,384	585	4,304	1.2%
accent	31,454	72	1,187	1.0%
trx	30,128	65	663	1.0%
gbp	28,426	59	356	0.9%
invsfoxepo01	24,254	723	4,411	0.8%
alv	21,877	155	1,293	0.7%
bqc	21,514	132	873	0.7%
mwp	18,982	99	760	0.6%
usrenalcare	17,294	73	855	0.5%
bqa	16,404	87	865	0.5%

Queries by dots in the name (ranked by number of queries, not dots)

All queries		
# dots	Queries	Cum %
2	1,734,024	54.9%
3	610,364	74.2%
5	227,779	81.4%
6	216,468	88.3%
1	140,928	92.7%
7	105,750	96.1%
4	86,286	98.8%
8	36,374	100.0%
9	637	100.0%
10	40	100.0%
11	21	100.0%
13	19	100.0%
0	12	100.0%
36	10	100.0%
12	1	100.0%

Two dots		
# dots	Queries	Cum %
2	1,668,354	100.0%

Underscore		
# dots	Queries	Cum %
6	206,353	29.4%
5	193,692	57.0%
7	104,900	72.0%
3	79,230	83.3%
4	50,767	90.5%
8	36,230	95.6%
2	29,941	99.9%
9	597	100.0%

Isatap or Wpad		
#	Queries	Cum %

dots		
2	65,670	50.7%
3	60,317	97.3%
4	3,189	99.7%
5	186	99.9%
6	92	99.9%
7	81	100.0%
8	1	100.0%

The rest		
# dots	Queries	Cum %
3	427,989	66.5%
1	140,926	88.4%
5	33,745	93.6%
4	29,838	98.3%
6	9,998	99.8%
7	766	100.0%
8	143	100.0%
10	40	100.0%
9	40	100.0%
11	21	100.0%
13	19	100.0%
0	12	100.0%
36	10	100.0%
12	1	100.0%

Queries by type of query

All queries		
QTYPE	**Queries**	**%**
A	1,778,960	56.3%
SOA	617,826	19.6%
SRV	547,345	17.3%
AAAA	138,116	4.4%
MX	50,782	1.6%
NS	8,340	0.3%
PTR	8,104	0.3%
TXT	4,755	0.2%
SPF	2,178	0.1%
ANY	1,830	0.1%
CNAME	416	0.0%
A6	51	0.0%
NAPTR	5	0.0%
DS	3	0.0%
HINFO	1	0.0%
RESERVED0	1	0.0%

Two dots		
QTYPE	**Queries**	**%**
A	1,225,926	73.5%
SOA	283,578	17.0%
AAAA	117,831	7.1%
MX	37,168	2.2%
ANY	1,524	0.1%
NS	1,325	0.1%
TXT	601	0.0%
SPF	243	0.0%
CNAME	107	0.0%
A6	51	0.0%

Underscore		
QTYPE	**Queries**	**%**
SRV	547,345	78.0%
SOA	95,351	13.6%
A	49,704	7.1%
PTR	7,981	1.1%
TXT	928	0.1%
CNAME	191	0.0%
AAAA	169	0.0%
MX	26	0.0%

NS	11	0.0%
ANY	4	0.0%

Isatap or Wpad		
QTYPE	**Queries**	**%**
A	129,194	99.7%
AAAA	342	0.3%

The rest		
QTYPE	**Queries**	**%**
A	359,699	55.9%
SOA	237,959	37.0%
AAAA	19,543	3.0%
MX	13,214	2.1%
NS	6,999	1.1%
TXT	3,651	0.6%
SPF	1,935	0.3%
ANY	297	0.0%
PTR	123	0.0%
CNAME	118	0.0%
NAPTR	5	0.0%
DS	3	0.0%
HINFO	1	0.0%
RESERVED0	1	0.0%