2014 Root DITL Data analysis and TLD popularity analysis

Kazunori Fujiwara, JPRS
<fujiwara@jprs.co.jp>

DNS-OARC 2014 Fall Workshop

Last Update: 2014/10/12 0205 (UTC)

DNS-OARC Root Datasets (1)

- "A Day in the Life of the Internet" (DITL) is a large-scale data collection project undertaken by CAIDA and DNS-OARC every year since 2006.
  - https://www.dns-oarc.net/oarc/data/ditl
  - 50 hours packet capture at root DNS servers and other DNS servers (48 hours are used by this analysis)
  - Source IP addresses of i.root-servers.net data are anonymized

DNS-OARC Root Datasets (2)

Year	Start (UTC)	End	List of root servers
2006	Jan 10 0000	Jan 12 0100	c,e,f,k (4/13)
2007	Jan 09 0000	Jan 11 0000	c,f,k,m (4/13)
2008	Mar 18 0000	Mar 20 0000	a,c,e,f,h,k,l,m (8/13)
2009	Mar 30 0000	Apr 02 0000	a,c,e,f,h,k,l,m (8/13), 72 hours
2010	Apr 14 0000	Apr 16 0000	a,b,c,d,e,f,g,h,i,j,k,l,m (12/13)
2011	Apr 12 1200	Apr 14 1200	a,c,d,e,f,h,j,k,l,m (10/13)
2012	Apr 17 1200	Apr 19 1200	a,c,e,f,h,j,k,l,m (9/13)
2013	May28 1200	May30 1200	a,c,d,e,f,h,j,k,l,m (10/13)
2014	Apr 15 1200	Apr 17 1200	a,c,e,f,h,j,k,m (8/13)

Differences between 2013 and 2014
48 hours data

Year	2013				2014			
Root servers	a,c,d,e,f,h,j,k,l,m (10/13)				a,c,e,f,h,j,k,m (8/13)			
	IP addresses		Number of Queries		IP addresses		Number of Queries	
Total	8,547,065		2.78E+10		10,087,711		3.21E+10	
RD0	6,081,035	71.15%	2.58E+10	92.79%	6,397,890	63.42%	2.99E+10	93.14%
EDNS0	3,572,804	41.80%	1.95E+10	70.17%	4,054,627	40.19%	2.38E+10	74.28%
DO=1	3,283,728	38.42%	1.89E+10	67.90%	3,751,076	37.18%	2.32E+10	72.32%
Update	228,633	2.67%	7.05E+07	0.25%	237,136	2.35%	1.26E+08	0.39%
Update Only	179,874	2.10%	3.99E+07	0.14%	182,447	1.81%	6.91E+05	0.22%
Non-exist	2,619,836	30.65%	1.17E+10	42.27%	2,563,956	25.42%	1.66E+10	51.90%
Exist	8,142,126	95.26%	1.52E+10	54.68%	9,575,391	94.92%	1.42E+10	44.32%
. NS	2,082,649	24.37%	6.47E+08	2.33%	2,220,978	22.02%	8.38E+08	2.61%
. Only	105,784	1.24%	6.25E+07	0.23%	200,267	1.99%	6.21E+07	0.19%
. DNSKEY	269,390	3.15%	8.50E+06	0.03%	521,733	5.17%	1.34E+07	0.04%

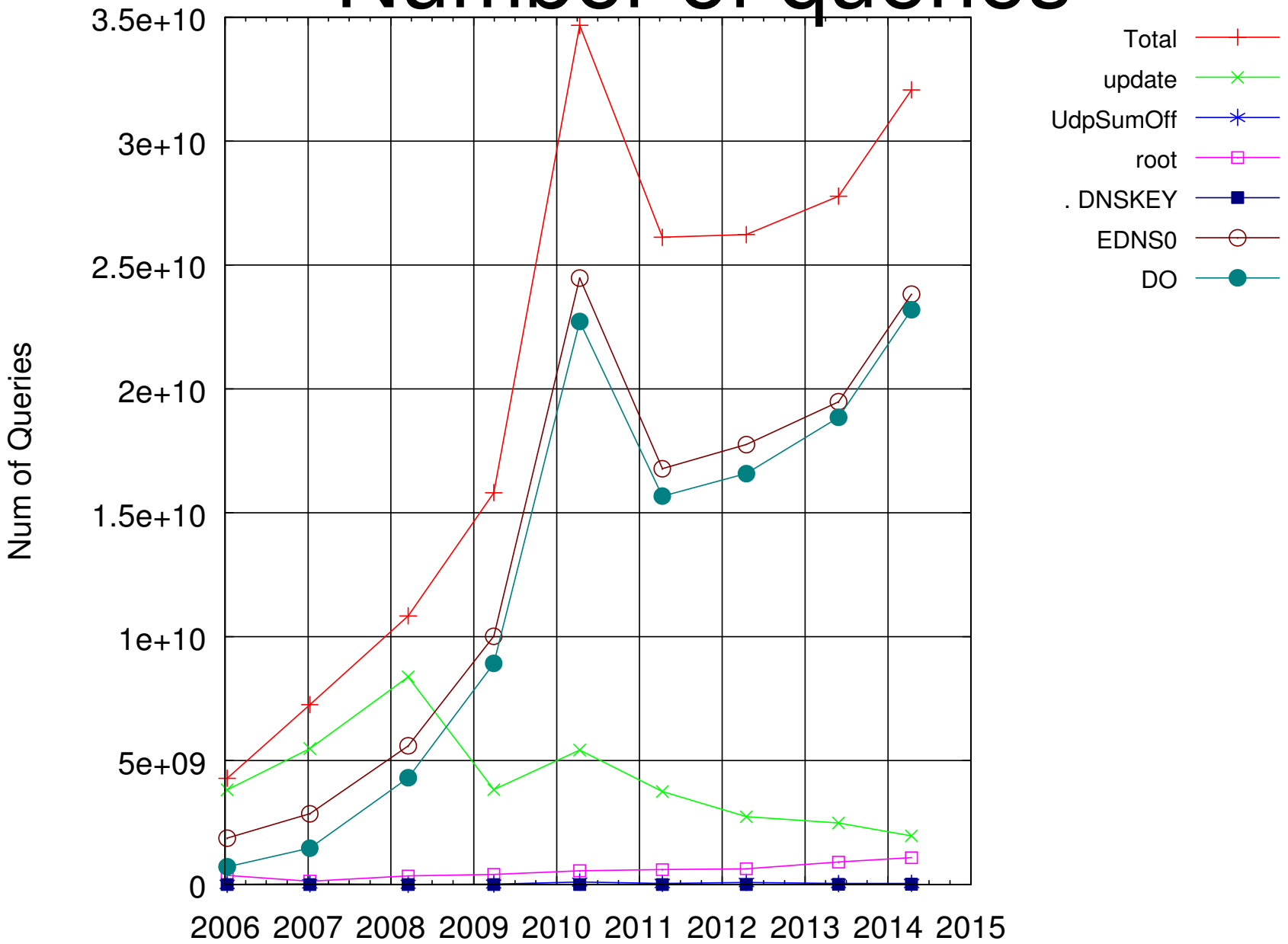
Differences between 2013 and 2014
48 hours data

Year	2013				2014			
Root servers	a,c,d,e,f,h,j,k,l,m (10/13)				a,c,e,f,h,j,k,m (8/13)			
	IP addresses		Number of Queries		IP addresses		Number of Queries	
Total	8,547,065		2.78E+10		10,087,711		3.21E+10	
RD0								93.14%
EDNS0								74.28%
DO=1								72.32%
Update	228,633	2.67%	7.05E+07	0.25%	237,136	2.35%	1.26E+08	0.39%
Update Only	179,874	2.10%	3.99E+07	0.14%	182,447	1.81%	6.91E+05	0.22%
Non-exist	2,619,836	30.65%	1.17E+10	42.27%	2,563,956	25.42%	1.66E+10	51.90%
Exist	8,142,126	95.26%	1.52E+10	54.68%	9,575,391	94.92%	1.42E+10	44.32%
. NS								%
. Only								%
. DNSKEY								%

2014 data derived from smaller number of roots (10 to 8)
However, both number of IP addresses and queries increased 18% and 15%

Increase of Non-existent TLD queries: 42.27% to 51.90%, 11.7B to 16.6B
However , number of IP addresses that send non-existent TLD queries decreased a little
Decrease of existent TLD queries: 54.68% to 44.32%, 15.2B to 14.2B

Number of queries

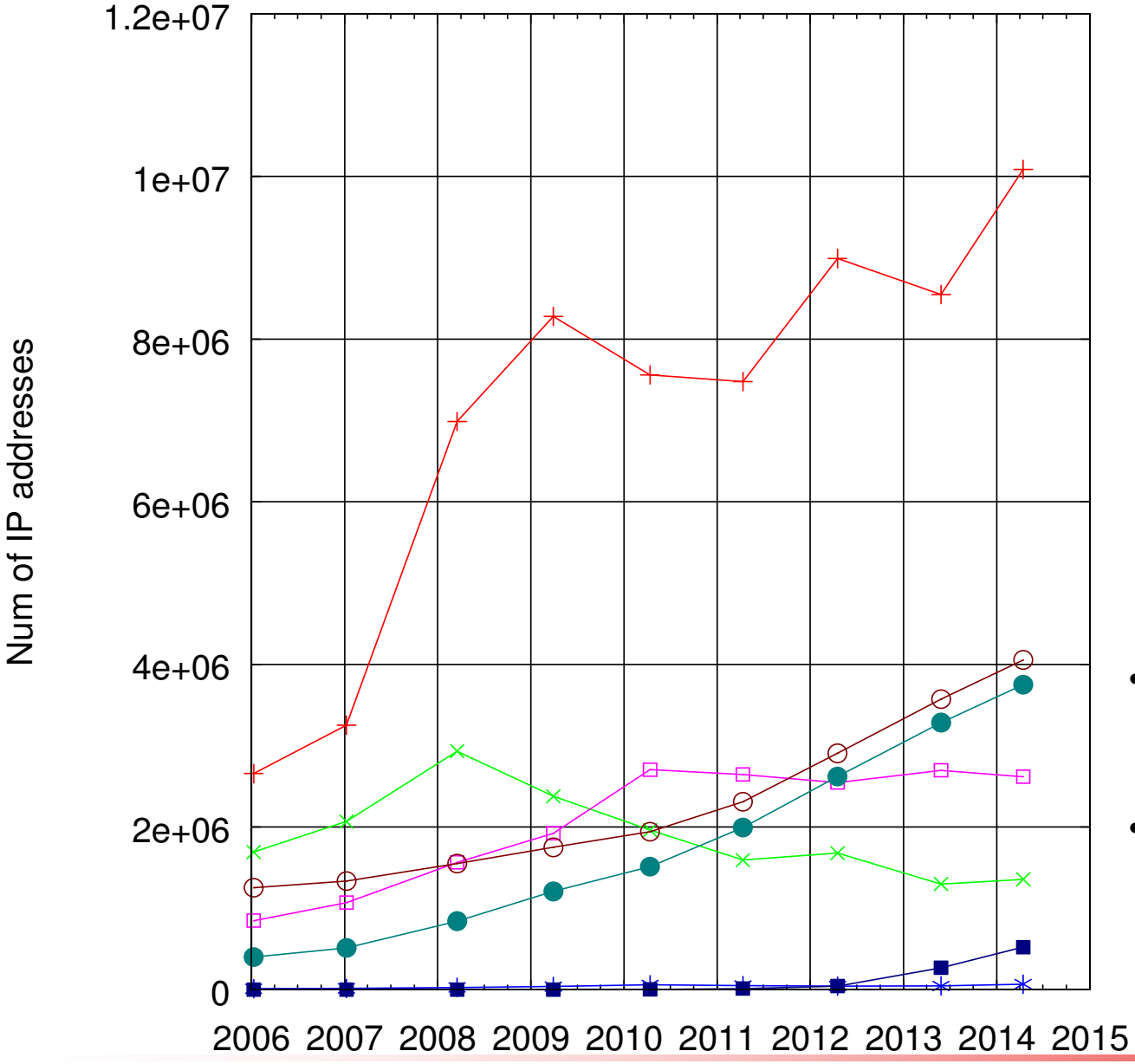


Number of IP addresses seen at root 48h

Year	2011		2012		2013		2014	
Data from	10 root		9 root		10 root		8 root	
Total	7,591,031		8,989,786		8,547,065		10,087,711	
RD0	5,846,612	77.0%	5,859,493	65.2%	6,081,035	71.1%	6,397,890	63.4%
EDNS0	2,340,543	30.8%	2,906,287	32.3%	3,572,804	41.8%	4,054,627	40.2%
DO=1	2,018,839	26.6%	2,621,660	29.2%	3,283,728	38.4%	3,751,076	37.2%
Update	105,131	1.4%	138,778	1.5%	228,633	2.7%	237,136	2.4%
Update Only	71,972	0.9%	99,902	1.1%	179,874	2.1%	182,447	1.8%
Non-exist	2,606,340	34.3%	2,641,072	29.4%	2,619,836	30.7%	2,563,956	25.4%
Exist	7,361,794	97.0%	8,697,606	96.7%	8,142,126	95.3%	9,575,391	94.9%
. NS	1,940,015	25.6%	1,871,995	20.8%	2,082,649	24.4%	2,220,978	22.0%
. Only	26,877	0.4%	36,920	0.4%	105,784	1.2%	200,267	2.0%
. DNSKEY (RD0)	14,092	0.2%	43,782	0.5%	269,390	3.2%	521,733	5.2%
. DNSKEY . Only	571	0.0%	2,828	0.0%	64,612	0.8%	146,752	1.5%

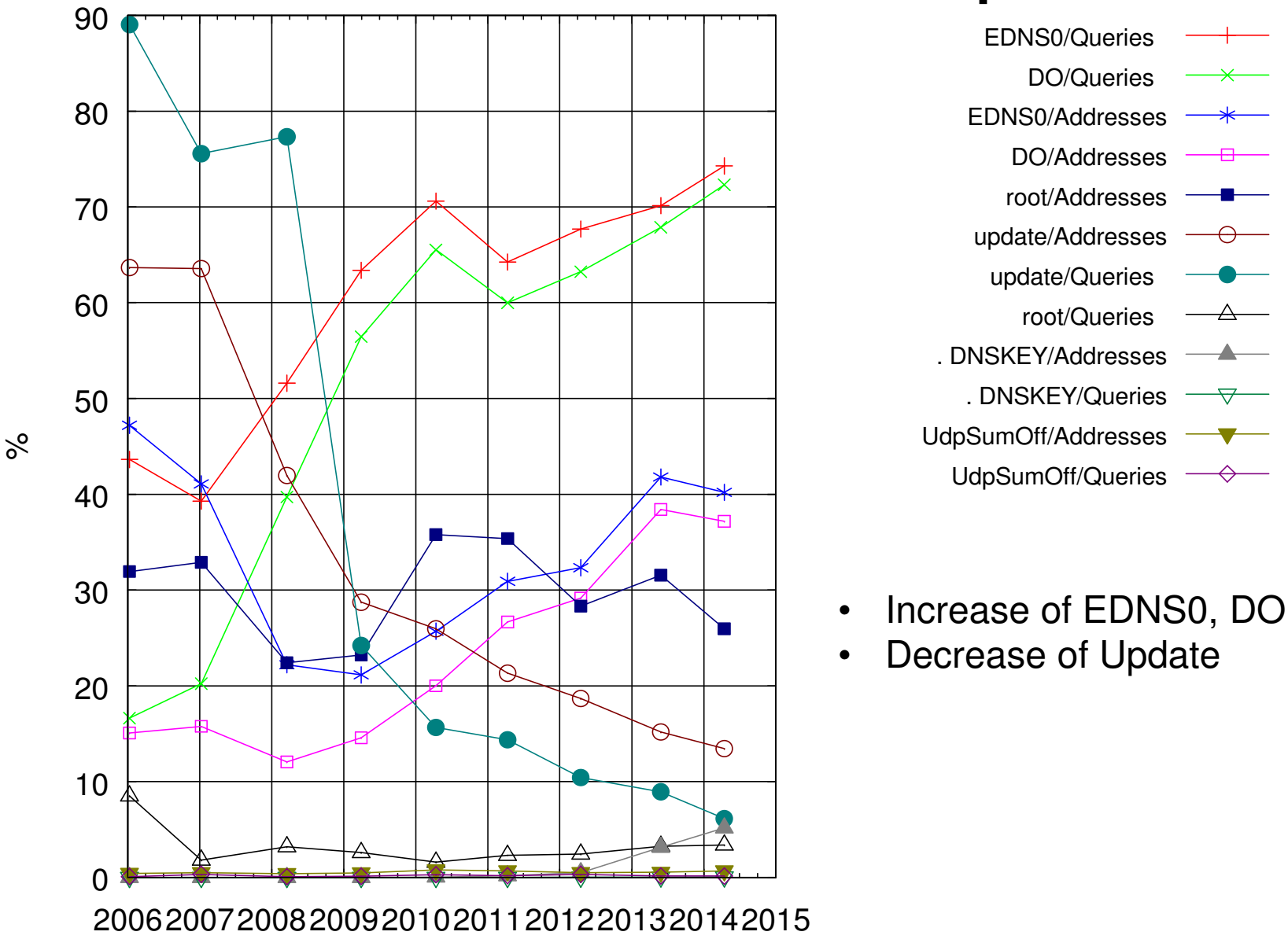
- EDNS0 and DO support is spreading gradually
  - (ratio decreased, number increased in 2014)
- Probable DNSSEC validators are still increasing (0.2% to 5.2%, 14,092 to 521,733)
- Some of them send "." queries only 571 to 146,752
  (RFC 5011 test ? Configuration only?)

Number of IP addresses



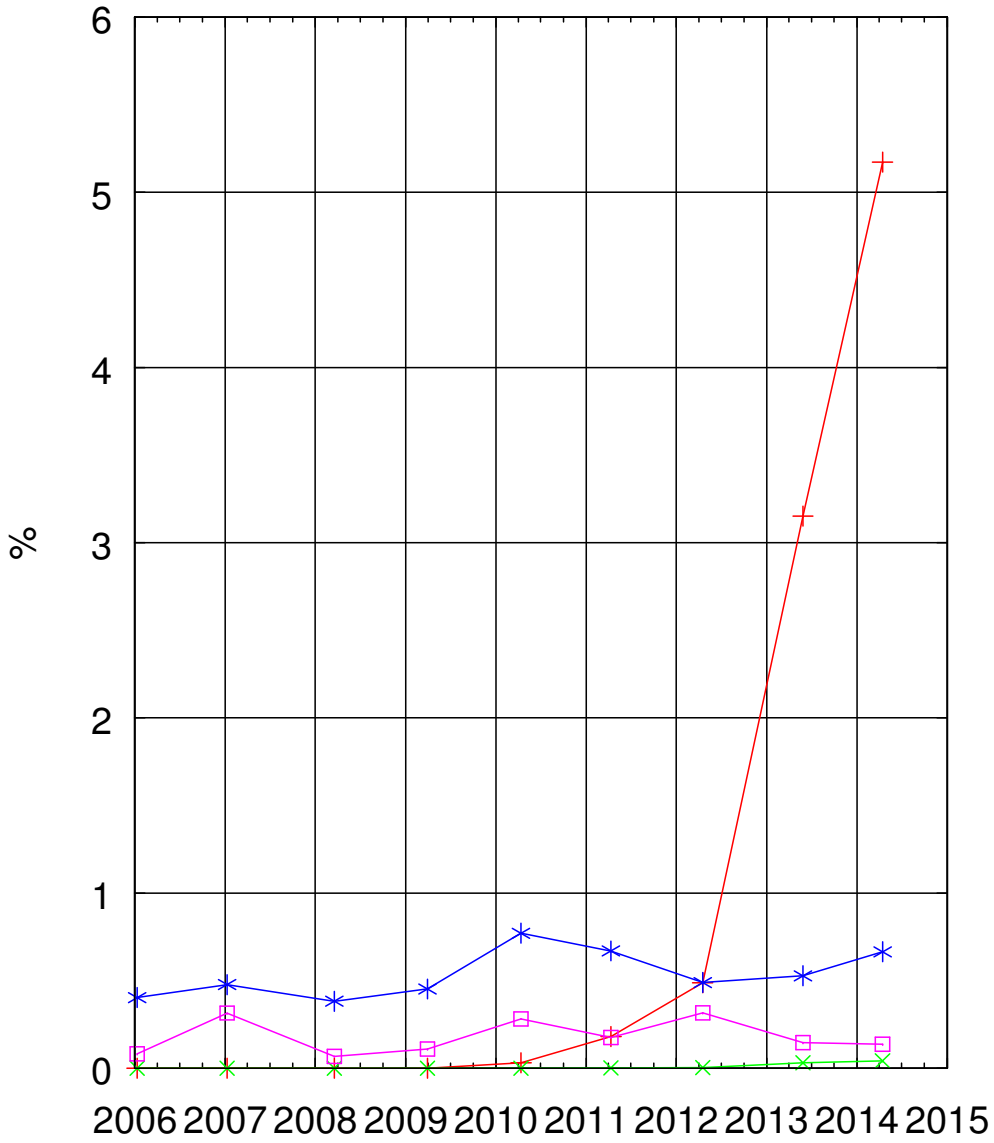
- Increase of IP addresses seen at root
- Increase of EDNS0, DO support, probable DNSSEC validators

Ratio of addresses/queries



- Increase of EDNS0, DO support
- Decrease of Update

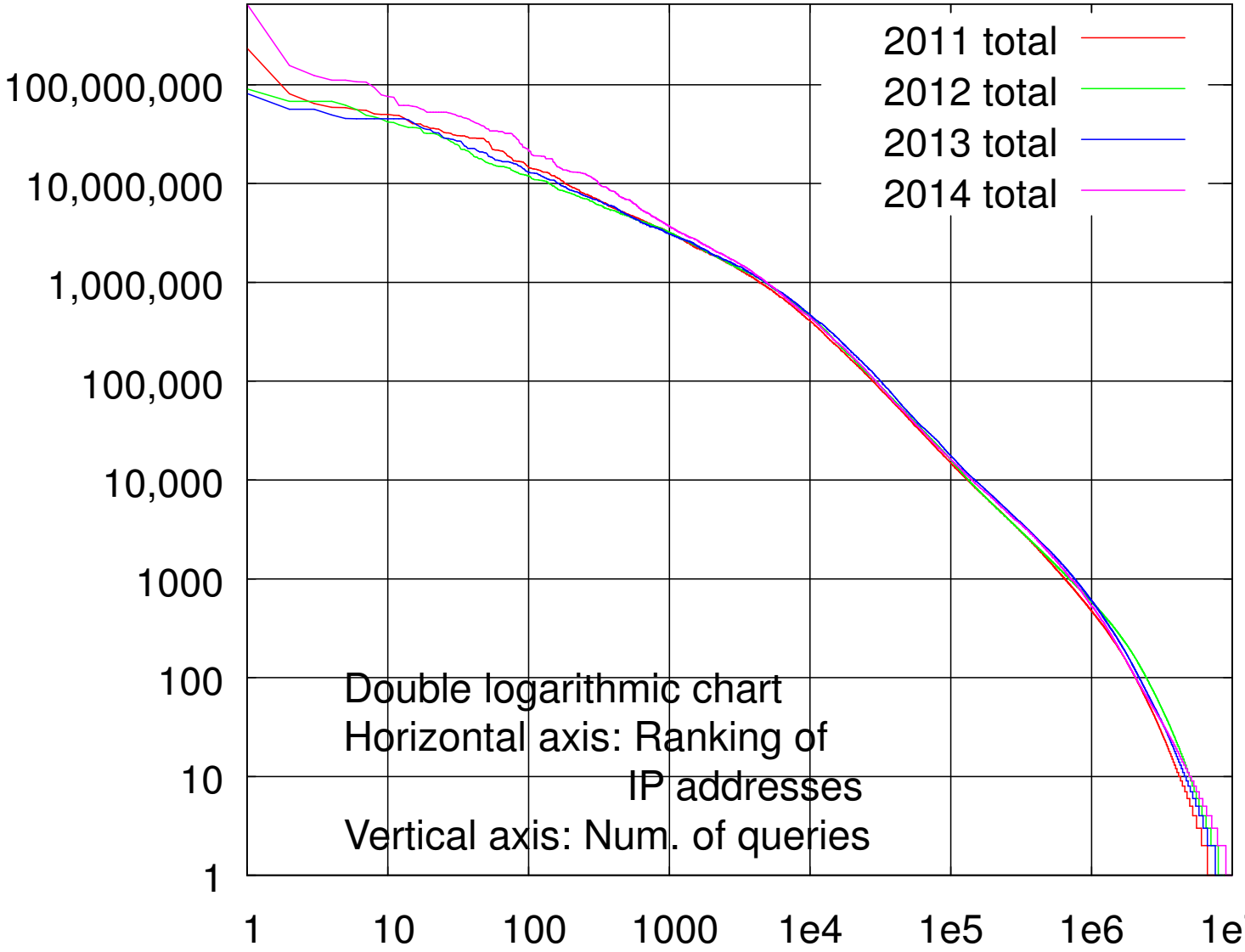
Ratio of addresses/queries (2)



Now 5% of IP addresses send
“.” DNSKEY query and
they may be DNSSEC validators

Analysis of JP query data shows
the same tendency as this (Root),
and about 6% of addresses send
JP DNSKEY queries now.

Number of queries sent from each address at root, 48 hours (recent 4 years)



Same tendency
However,
addresses of
higher rank
sent more
queries from
The previous year

No.1 address
tent 654 million
queries to root
(3786 pps ave !!)

No 2 to 7
sent 100 million
to 150 million
queries to root

Result of 2014 DITL data analysis

- Both number of IP addresses and queries increased (18% and 15) in spite of data source root servers decreased (10 to 8)
- Number of IP addresses which support EDNS0 and DNSSEC increased (about 500,000), however ratio decreased a little
- Number of probable DNSSEC validators is still increasing and now 5% of IP addresses may be DNSSEC validators
- There is a strange IP address which sent over 600,000,000 queries within 48 hours

TLD popularity analysis

TLD popularity ?

- Each query to root contains TLD name
- Assumption
  - If an end user access to a website in a TLD, a full-resolver send queries which contain the TLD to Root
- Method
  - For each IP address, extract TLDs from queries which the address sent to root
    - It is recorded in tldbitmap
  - For each TLD, count number of IP addresses which sent the TLD queries

TLD list

- Delegated TLDs from root zone
  - Delegated date is important to classify TLD existence (in previous analysis)
- Some popular non-existent TLDs
  - bind (version.bind)
  - server (id.server)
  - local
  - localhost localdomain loghost localnet internal
  - i2p exit zkey gnu onion bit
    - from draft-grothoff-iesg-special-use-p2p-names
  - home belkin alt flets

TLD ranking (query volume), 2014, 48h

rank	TLD	Num of queries	Ratio of queries	Num of IP addrs	Ratio of IP addrs
1	com	6.3E+09	19.6%	7,707,003	76.4%
2	net	2.92E+09	9.1%	4,714,908	46.7%
3	local	1.84E+09	5.7%	958,831	9.5%
4	home	7.67E+08	2.4%	258,256	2.6%
5	org	5.75E+08	1.8%	3,602,061	35.7%
6	cn	5.09E+08	1.6%	1,765,917	17.5%
7	arpa	4.96E+08	1.5%	2,239,253	22.2%
8	internal	4.26E+08	1.3%	95,269	0.9%
9	localdomain	3.97E+08	1.2%	228,244	2.3%
10	localhost	3.18E+08	1.0%	85,721	0.8%
11	ru	2.36E+08	0.7%	1,593,528	15.8%
12	belkin	1.95E+08	0.6%	105,200	1.0%
13	uk	1.86E+08	0.6%	2,366,576	23.5%
14	de	1.75E+08	0.5%	2,230,824	22.1%
15	info	1.39E+08	0.4%	2,261,243	22.4%
16	jp	1.38E+08	0.4%	1,437,134	14.2%

TLD ranking (query volume), 2014, 48h

rank	TLD	Num of queries	Ratio of queries	Num of IP addrs	Ratio of IP addrs
1	com	6.3E+09	19.6%	7,707,003	76.4%
2	net	2.92E+09	9.1%	4,714,908	46.7%
3	local	1.84E+09	5.7%	958,831	9.5%
4	home	7.67E+08	2.4%	258,256	2.6%
5	org	5.75E+08	1.8%	3,602,061	35.7%
6	cn	5.09E+08	1.6%	1,765,917	17.5%
7	arpa	4.96E+08	1.5%	2,239,253	22.2%
8	internal	4.26E+08	1.3%	95,269	0.9%
9	localdomain	3.97E+08	1.2%	228,244	2.3%
10	localhost	3.18E+08	1.0%	85,721	0.8%
11	ru	2.36E+08	0.7%	1,593,528	15.8%
12	belkin	1.95E+08	0.6%	105,200	1.0%
					5%
					1%
					4%
					2%

com, net are very popular (28.7%)
Root servers receives many non delegated popular TLD queries
Their query volumes are higher than delegated TLDs (because of TTL)
local is leaking well

TLD ranking (IP addrs), 2014, 48h

rank	TLD	Num of IP addrs	Ratio of addrs	Ratio of Queries
1	com	7,707,003	76.4%	19.6%
2	net	4,714,908	46.7%	9.1%
3	org	3,602,061	35.7%	1.8%
4	uk	2,366,576	23.5%	0.6%
5	info	2,261,243	22.4%	0.4%
6	arpa	2,239,253	22.2%	1.5%
7	de	2,230,824	22.1%	0.5%
8	cn	1,765,917	17.5%	1.6%
9	ru	1,593,528	15.8%	0.7%
10	fr	1,542,618	15.3%	0.2%
11	biz	1,520,356	15.1%	0.2%
12	eu	1,453,687	14.4%	0.2%
13	jp	1,437,134	14.2%	0.4%
14	nl	1,414,254	14.0%	0.2%
15	us	1,394,847	13.8%	0.2%

TLD ranking (IP addrs), 2014, 48h

rank	TLD	Num of IP addrs	Ratio of addrs	Ratio of Queries
				.6%
				.1%
				.8%
				.6%
				.4%
				.5%
				.5%
				.6%
				.7%
				.2%
				.2%
				.2%
13	jp	1,437,134	14.2%	0.4%
14	nl	1,414,254	14.0%	0.2%
15	us	1,394,847	13.8%	0.2%

Comparison with JP data

- JPRS collects packet captures of all JP DNS servers, around the same timing as DNS-OARC DITL 2014
- There were 2,129,261 IP addresses that sent *.JP queries to JP DNS servers, at the same timing of 2014 DITL
- Root data shows that there were 1,437,134 IP addresses which interested JP
- They are almost the same value because the DITL dataset does not cover all root DNS servers (8 of 13)

TLD ranking (IP addrs), 2014, 48h

- No 1 to 120

1 to 10: com net org uk info arpa de cn ru fr
11 to 20: biz eu jp nl us tv it co br au
21 to 30: se ca edu in me pl kr es gov fi
31 to 40: at ch cz mx be local io tw dk ar
41 to 50: tr hk cc ua no ro sg za nz gr
51 to 60: ly il ms id my pt vn la ie ws
61 to 70: hu sk cl fm mobi lt th to bg li
71 to 80: st is gl am by ae lv pe ph kz
81 to 90: hr ee su lu mil si rs name asia
91 to 100: pk ir nu im uy pro sa tk do ma pw
101 to 110: int ve bz md ec cr gs sh tn re
111 to 120: ag home so lk ad az travel tl bo jobs

Query source address analysis

- Used maxmind GeoLite Country database
  - Added some rule (using whois, traceroute)
- We can analyze TLD popularity by each county

Query distribution by countries

rank	Country	Number of IP addrs	Ratio of IP address	Number of Queries	Ratio of Queries
1	CN	2,597,365	25.7%	1.64E+09	23.5%
2	US	1,772,034	17.6%	9.98E+08	14.3%
3	DE	813,470	8.1%	5.97E+08	8.6%
4	FR	394,450	3.9%	3.05E+08	4.4%
5	GB	292,809	2.9%	1.71E+08	2.5%
6	RU	291,891	2.9%	2.54E+08	3.6%
7	BR	271,171	2.7%	1.48E+08	2.1%
8	JP	235,017	2.3%	1.69E+08	2.4%
9	IT	215,775	2.1%	1.27E+08	1.8%
10	CA	210,507	2.1%	1.14E+08	1.6%
11	IN	188,107	1.9%	1.09E+08	1.6%
12	AU	172,849	1.7%	7.48E+07	1.1%

Two ideas of TLD usage ratio

Note: this analysis is experimental
because cache removes real usage

1. Ratio of IP addresses which interests TLD
   - Normalized (total 100%) number of IP addresses that interest TLD
   - Density of each address is not the same
   - S1(TLD) = $\frac{NumberOfIPaddresses(TLD)}{\sum_{all\ TLDs} NumberOfIPaddress(TLD)}$
2. Sum of usage share of each IP addresses
   - Assumption: Density of each address is the same
   - If an address sends multiple (n) TLD queries, the AddressShare(addr,TLD) becomes 1/n.
   - S2(TLD) = $\frac{\sum_{All\_IP\_addresses} AddressShare(addr,TLD)}{Number\_of\_IP\_addresses}$
   - It becomes large if an address sends a TLD query only

TLD usage (seen at root, 2014)

rank	TLD	Number of IP addrs	Ratio of IP addrs	S1: TLD Interests	S2: Usage ratio
1	com	7,707,003	76.4%	6.4%	39.8%
2	net	4,714,908	46.7%	3.9%	11.8%
3	org	3,602,061	35.7%	3.0%	5.6%
4	uk	2,366,576	23.5%	2.0%	2.1%
5	info	2,261,243	22.4%	1.9%	1.9%
6	arpa	2,239,253	22.2%	1.9%	3.7%
7	de	2,230,824	22.1%	1.9%	2.3%
8	cn	1,765,917	17.5%	1.5%	1.9%
9	ru	1,593,528	15.8%	1.3%	1.2%
10	fr	1,542,618	15.3%	1.3%	0.9%
11	biz	1,520,356	15.1%	1.3%	1.0%
12	eu	1,453,687	14.4%	1.2%	0.7%
13	jp	1,437,134	14.2%	1.2%	0.9%
14	nl	1,414,254	14.0%	1.2%	0.7%

TLD usage in Japan (2014)

All IP addresses(10,087,711)			
Rank	TLD	S1:TLD Interest	S2: Usage ratio
1	com	6.4%	39.8%
2	net	3.9%	11.8%
3	org	3.0%	5.6%
4	uk	2.0%	2.1%
5	info	1.9%	1.9%
6	arpa	1.9%	3.7%
7	de	1.9%	2.3%
8	cn	1.5%	1.9%
9	ru	1.3%	1.2%
10	fr	1.3%	0.9%
11	biz	1.3%	1.0%
12	eu	1.2%	0.7%
13	jp	1.2%	0.9%
14	nl	1.2%	0.7%
15	us	1.2%	0.7%

IP addresses in Japan(235,017)			
Rank	TLD	S1: TLD Interest	S2: Usage ratio
1	com	4.3%	19.4%
2	net	3.9%	12.4%
3	jp	3.5%	11.6%
4	org	3.0%	6.6%
5	info	2.3%	3.4%
6	uk	2.1%	2.9%
7	arpa	2.0%	4.7%
8	cn	1.9%	2.2%
9	de	1.6%	1.6%
10	fr	1.3%	1.2%
11	biz	1.3%	1.1%
12	br	1.3%	1.1%
13	kr	1.2%	0.8%
14	ru	1.2%	0.8%
15	nl	1.1%	0.7%

TLD usage in each country (1)

IP addresses in CN			
Rank	TLD	S1 %	S2 %
1	com	42.59	89.21
2	cn	4.41	3.10
3	net	3.89	2.31
4	org	2.19	0.82

IP addresses in US			
Rank	TLD	S1 %	S2 %
1	com	3.95	21.06
2	net	3.78	20.78
3	org	2.73	7.35
4	uk	1.95	2.87
5	arpa	1.85	5.19

IP addresses in DE			
Rank	TLD	S1 %	S2 %
1	com	5.80	22.32
2	net	4.81	13.92
3	de	4.01	9.38
4	org	3.69	7.27

IP addresses in FR			
Rank	TLD	S1 %	S2 %
1	com	5.21	17.32
2	net	4.91	12.88
3	org	3.54	6.99
4	fr	3.22	6.40

TLD usage in each country (2)

IP addresses in RU			
Rank	TLD	S1 %	S2 %
1	com	3.38	16.35
2	net	2.97	10.89
3	ru	2.87	11.01
4	org	2.32	5.95

IP addresses in GB			
Rank	TLD	S1 %	S2 %
1	com	4.74	21.94
2	net	4.24	17.85
3	org	3.06	7.35
4	uk	3.05	7.73

IP addresses in BR			
Rank	TLD	S1 %	S2 %
1	com	4.10	17.46
2	net	3.60	12.19
3	br	3.00	8.57
4	org	2.91	6.77

IP addresses in IT			
Rank	TLD	S1 %	S2 %
1	com	5.09	22.29
2	net	4.22	12.83
3	org	3.37	7.34
4	it	3.33	8.58

Result of TLD popularity

- Very popular gTLDs are com, net, org, arpa, info
- uk, cn, de, ru are very popular ccTLDs

- JP TLD is very popular in Japan
  - However, com, net are more popular than jp
- ccTLD is very popular in many counties except US
  - The order of other TLDs is similar to all IP address data
- These results suit feelings

Conclusion

- Analyzed DITL 2014 data briefly
- Tried to show TLD usage seen at Root

Acknowledgements

- DNS-OARC as the data source of Root dataset