CDAR

Continuous Data-driven Analysis of Root Stability

Progress report

DNS WG Ripe 73

Jaap Akkerhuis (NLnet Labs)

DISCLAIMER: All data subject to change

CDAR Study

- **Objective**
  Analyze the technical impact of the introduction of new gTLDs on the stability & security of the root server system

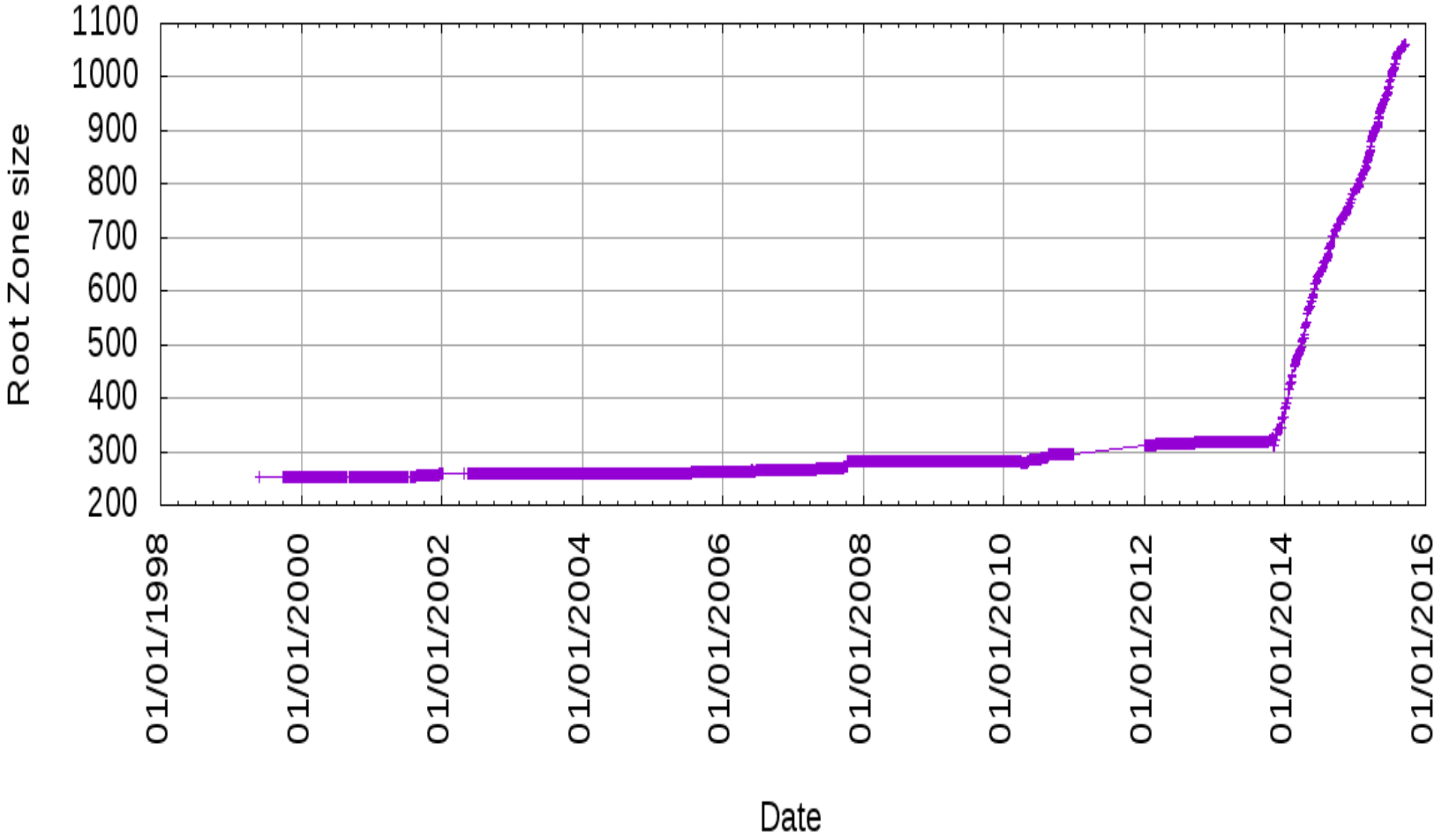
- **Approach**
  - Data-driven, using wide variety of DNS data
    RSSAC002, DITL '13, '14, `15 (, '16), RSO's PCAP and DSC data, ATLAS / DNSMON, Zone File Repository, gTLD Registry reports, specific tools and public data sources
  - Interaction with the broader tech community
    ICANN and advisory committees, RSOs, DNS-OARC, IEPG/IETF
  - Using and expanding previous studies of RSS behavior
    DITL papers, L-Root scaling report, Name Collision report, …
    DNS Health reports and DNS threat analysis papers

Breaking news!

The Draft Report has just been send to ICANN

Public comment period will start soon

Hockey stick grow



NLnet Labs SIDN LABS TNO

Stability of the Root Server System

- **Stability levels**

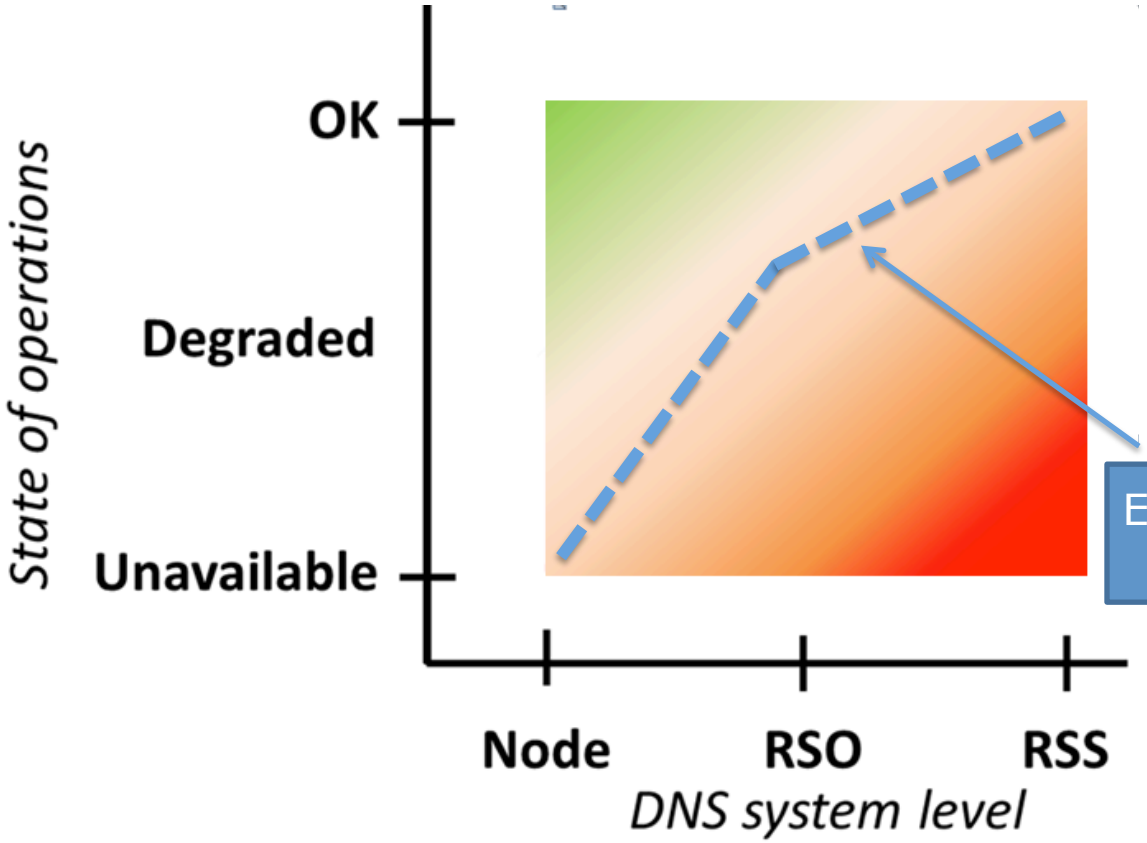
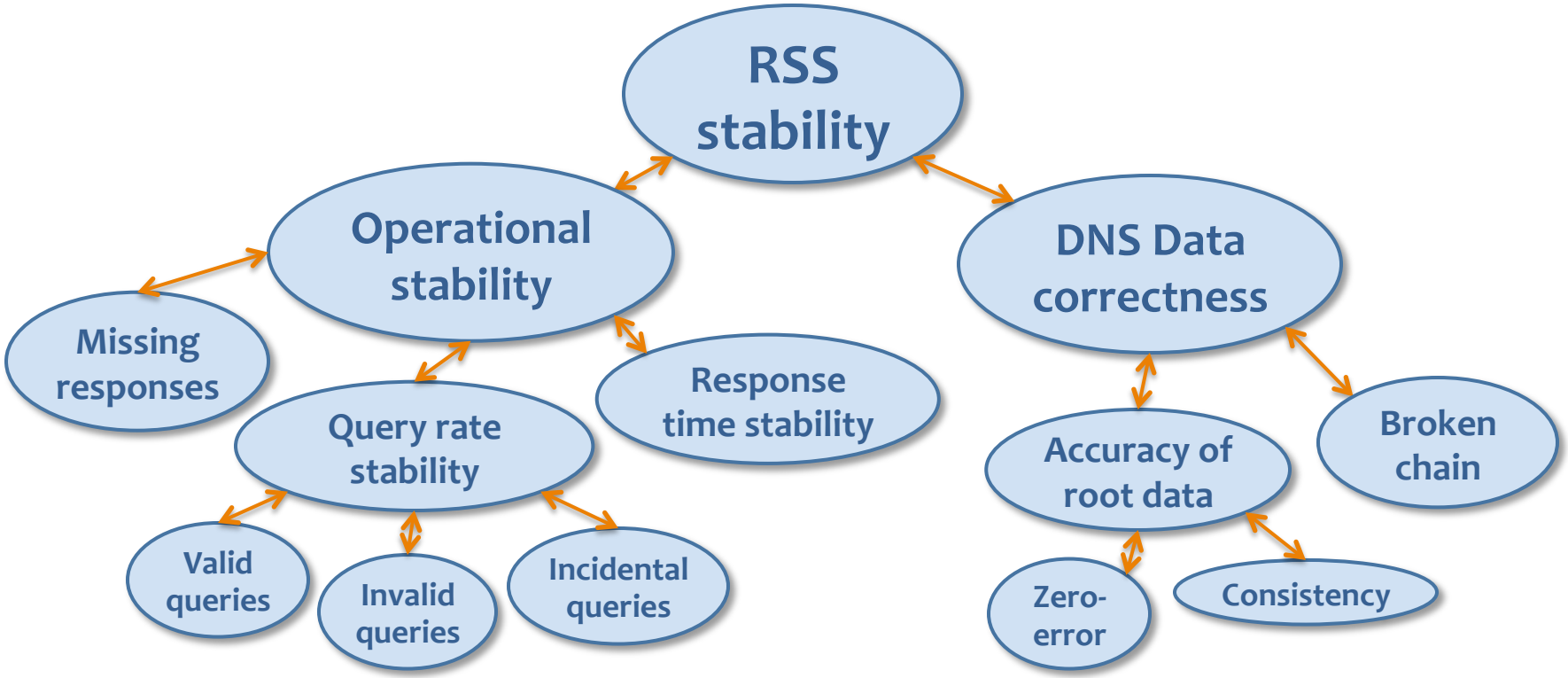


Illustration of the distinction between macroscopic and microscopic scale for RSS stability

Example of RSS stability level: Nov./Dec. `15 attack traffic

Possible Impact of New gTLD

- **Analysis of a range of RSS stability indicators**
  - As observed from the 'outside'



Precaution: Limitations of Data Used

- Limited data accuracy
  - Measurement breakdowns don't always add up to the total
  - Different data collection methods lead to different results
- Incomplete
  - Collected data sets are not always complete
  - Not all relevant partners contribute to RSSAC002, DITL etc
  - History of some data sets is relatively short
- Lack of standardized data format
  - Different sources use different data formats and collection
  - Is improving with RSSAC002 and DITL

CDAR Analyses

			Data sources									
			DNS OARC	RSO's	DNS OARC	H-Root	RIPE	ICANN	NLnet Labs	RSO's	ICANN	Public web
Hyp. Group	**Hyp. ID**	**Hypothesis**	**ZFR**	**RSSAC002**	**DITL**	**Renumb. Data**	**DNSMON**	**new gTLD monitoring**	**DNSSEC broken chain**	**DSC**	**Registry reports**	**#domain/TLD**
query rates (impact of new gTLDs)	1.1	TTL value characteristics for New gTLDs are comparable to TTL values of other TLDs	√									
	1.2	Cache hit rates for New gTLDs are comparable to cache hit rates for other TLDs			√	√				√	√	√
	1.2a	*Fraction of identical queries for New gTLDs is not significantly higher than for other TLDs*										
	1.2b	*Fraction of repeated queries for New gTLDs is not significantly higher than for other TLDs*										
	1.2c	*Fraction of referral-not-cached for New gTLDs is not significantly higher than for other TLDs*										
	1.3	Increasing the number of TLDs does not significantly increase the query rate to the root	√	√								
	1.4	The ratio between #domains in a TLD and query rate to the DNS root are comparable for New gTLDs and other TLDs		√	√	√				√	√	√
	1.5	When a New gTLD is first delegated in the RZF this has non-significant impact on the query rate to the Root in the period immediately after the delegation	√			√						
	1.6	The New gTLD data in the RZF does not change much more frequent than for other TLDs	√									
	1.7	When New gTLD data in the RZF changes this has non-significant impact on the query rate to the Root	√			√						
	1.8	The introduction of New gTLDs has non-significant impact on the amount of bogus traffic ending up at the Root										
	1.9	IoT (explosion of number of deviced connected to the internet) and Mobile internet traffic does not increase due to New gTLDs										
	1.10	Many domains in New gTLDs are redirected to regular TLDs, especially for dot-brands										
	1.10a	New gTLD traffic redirected to other/existing TLDs hardly generates extra traffic to the DNS root										
	1.11	The number of lame delegations per TLD is correlated with the query rate per TLD										
response characteristics (impact of new gTLDs)	2.1	Response size statistics (average, maximum, percentile values) for responses by the Root DNS for New gTLD queries are not significantly larger than the sizes for other TLDs			x							
	2.2	Response type distribution characteristics from the Root DNS for New gTLD queries are not significantly different from the characteristics for other TLDs			x							
	2.3	Ratio of TCP/UDP queries will be higher for New gTLD than for other TLDs (due to DNSSEC)			√							
	2.4	Query type distribution characteristics for new gTLD queries to the root are not significantly different from the characteristics for other TLDs			√							
RTT / data availability (impact of new gTLDs)	3.1	RTT is not significantly affected immediately after the delegation of New gTLDs to the RZF	√				√					
	3.2	The fraction of queries not answered is not significantly affected immediately after the delegation of New gTLDS to the RZF	√				√					
	3.3	The RTT for new gTLDs is not significantly larger than the RTT for other TLDs					√	√				
	3.4	The fraction of queries not answered for New gTLDs is not significantly larger than the fraction of queries not answered for other TLDs					√					
	3.5	There is no correlation between the RTT and the total number of TLDs	√				√					
Data correctness (impact of new gTLDs)	4.1	DNSSEC is used more often for New gTLDs than for TLDs	√									
	4.2	DNSSEC validation errors (broken chain) does not occur more frequently for New gTLDs, than for other TLDs							√			
	4.3	The number of errors in the RZF is not significantly increased after introduction of New gTLDs	√									
	4.4	New gTLDs zone files do not contain significantly more errors than zone files for other TLDs										
stability of data center facilities (impact	5.1	The majority of New gTLDs make use of name servers owned by a small number of very experienced back-end registry providers										
	5.1a	*And therefore name servers used by New gTLDs are in general very stable and secure*										

CDAR Analyses Presented Today

What is impact of increased number of TLDs on the query rate to the root?

Is the ratio between #domains in a TLD and the query rate to the root comparable for New gTLDs and other TLDs?

Are cache hit rates for New gTLDs comparable to cache hit rates for other TLDs?

What is the impact of a new gTLD's initial delegation in the RZF on the query rate to the Root (in the delegation period)?

What is the impact of a new gTLD's initial delegation on the RTT?

Do DNSSEC validation errors (broken chain) occur more frequently for New gTLDs, than for other TLDs?

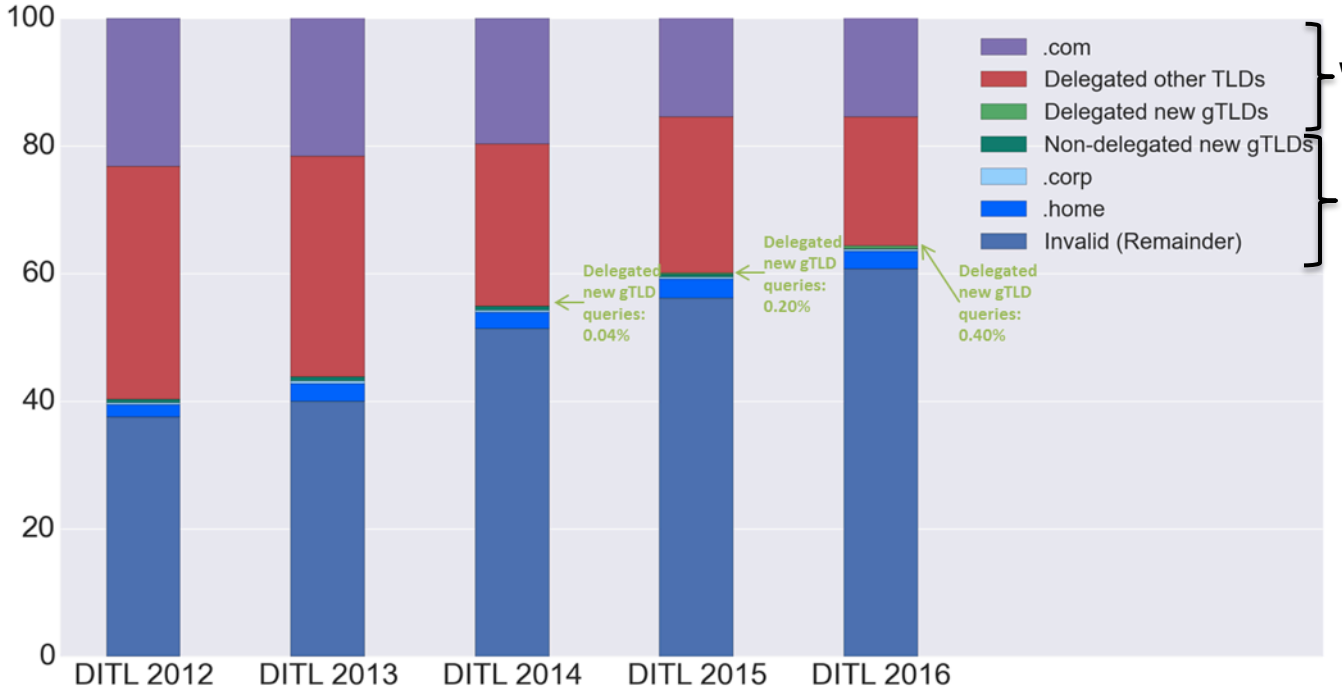
What is the behaviour of resolvers with validation errors?

Does the query type distribution for queries to new gTLDs differ from the query type distribution for queries to other TLDs?

Is there a geographic affinity for geographic new gTLDs?

New gTLD Queries to the Root

- ***The percentage of queries*** *to New gTLDs has increased over time, but is still very low compared to other queries*



- Using DITL data offers possibility to relate results to period prior to new gTLD program

A note on historic "invalid queries":
- F-root analysis Jan2001: %Invalid = 20%
- F-root analysis Oct2002: %Invalid = 19,6%
- DITL Mar2009: %Invalid ≈ 30+ %

Rule of Thumb for Valid Query Rates

- Data shows that a TLD's valid query rate to the root is 'bound' by the number of domains in the TLD
  - Example from DITL'15, K-root data:

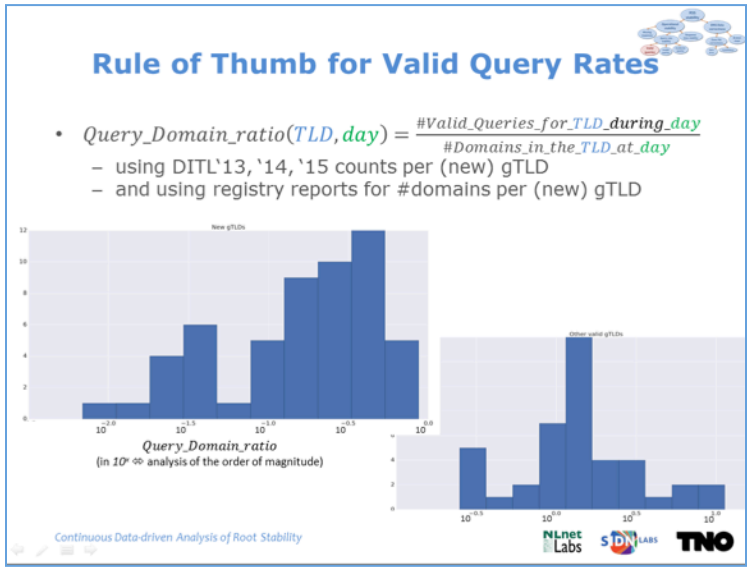
TLD	Nr. of queries / TLD	Nr. of domains / TLD	Query/domain ratio
.com	779.171.677	120.585.440	6,46 E+00
.org	91.095.714	10.569.583	8,62 E+00
.cn	51.949.760	11.678.026	4,45 E+00
.br	15.696.021	3.568.492	4,40 E+00
.club	651.082	202.519	3,21 E+00
.xyz	420.885	842.340	5,00 E-01

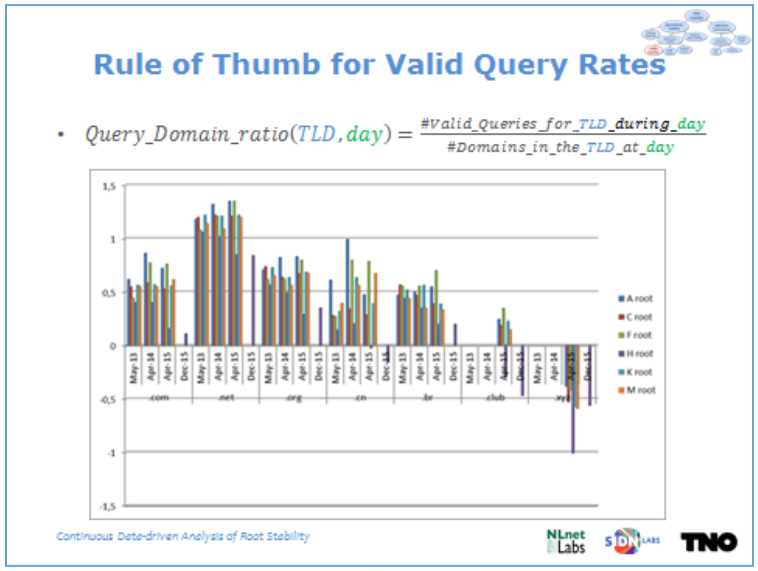
- More in general, this ratio rarely exceeds 10
  - For any TLDs in recent H-Root data sets
  - For new gTLDs the ratio is lower than for other TLDs

NLnet Labs SIDN LABS TNO

Rule of Thumb for Valid Query Rates

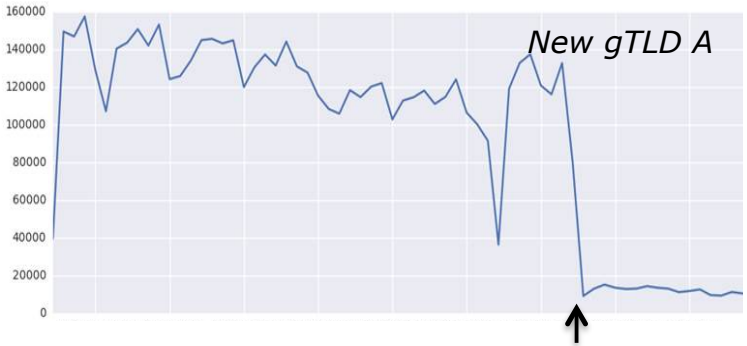
- A statistical rule of thumb
  - Intuition: #valid TLD queries to the root is 'bound' by the TLDs 'popularity' (in terms of #domains)
  - No DNS rational found for this observation
- This observation might provide an easily verifiable bound on valid queries to the root

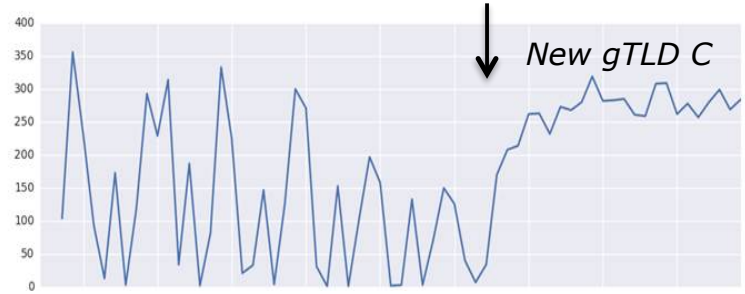


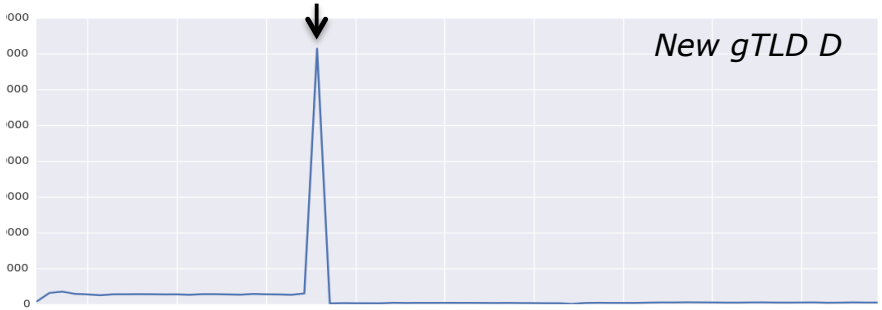


Impact of Initial Delegation (Query Rate)

The volume of root traffic for a new gTLD often decreases significantly after delegation (gTLDs A and B), but sometimes also increases (gTLD C) or increases temporarily (gTLD D)



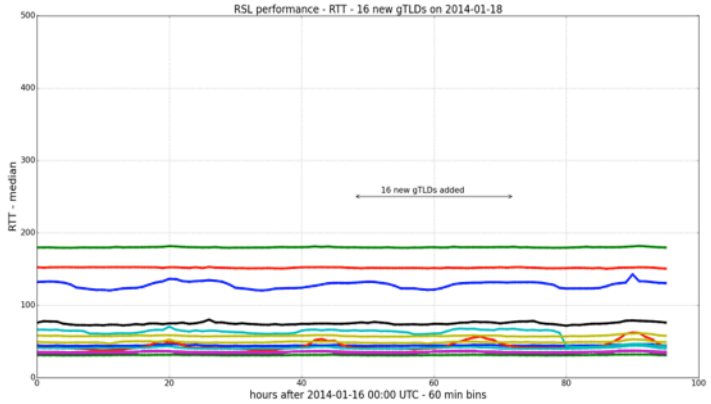




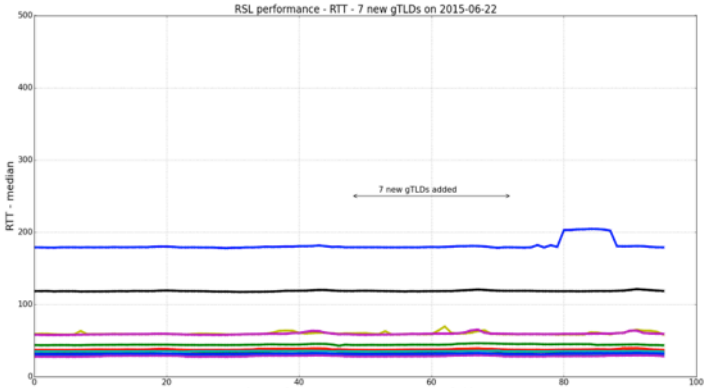
Impact of Initial Delegation (RTT)

RTT is not significantly affected after delegation of New gTLDs to the RZF

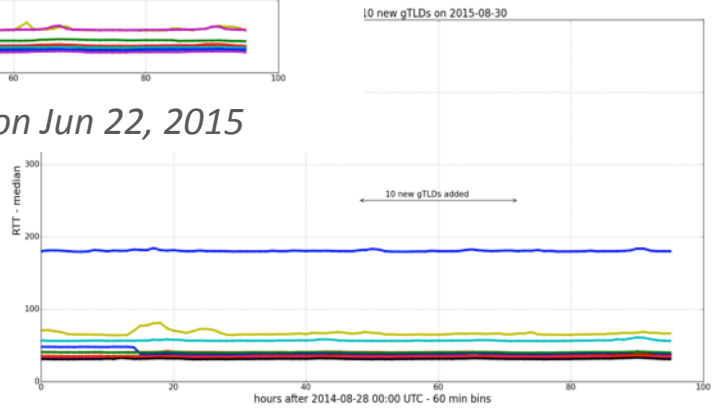
- Considering 22 days in which 7 or more new gTLDs were delegated
- Using RTT measurements from Atlas RIPE to all root servers



16 new gTLDs delegated on Jan 18, 2014



7 new gTLDs delegated on Jun 22, 2015



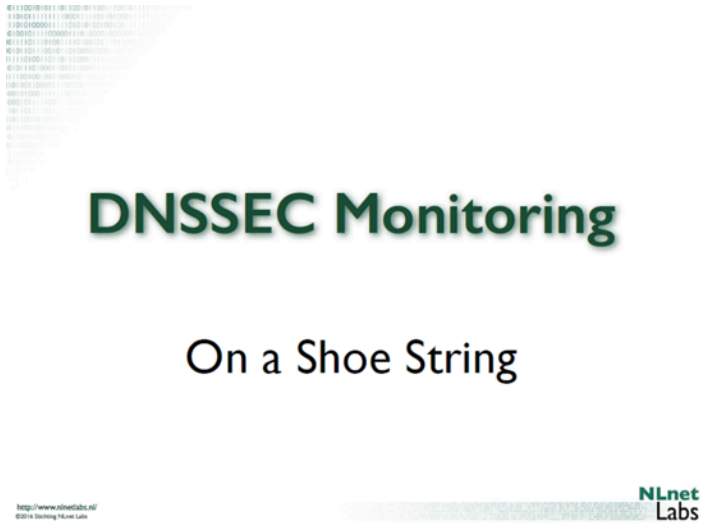
10 new gTLDs delegated on Aug 30, 2015

In general changes in the RTT before and after delegation are **minor**, both up and down

NLnet Labs SIDN LABS TNO

Impact on DNS Data Correctness

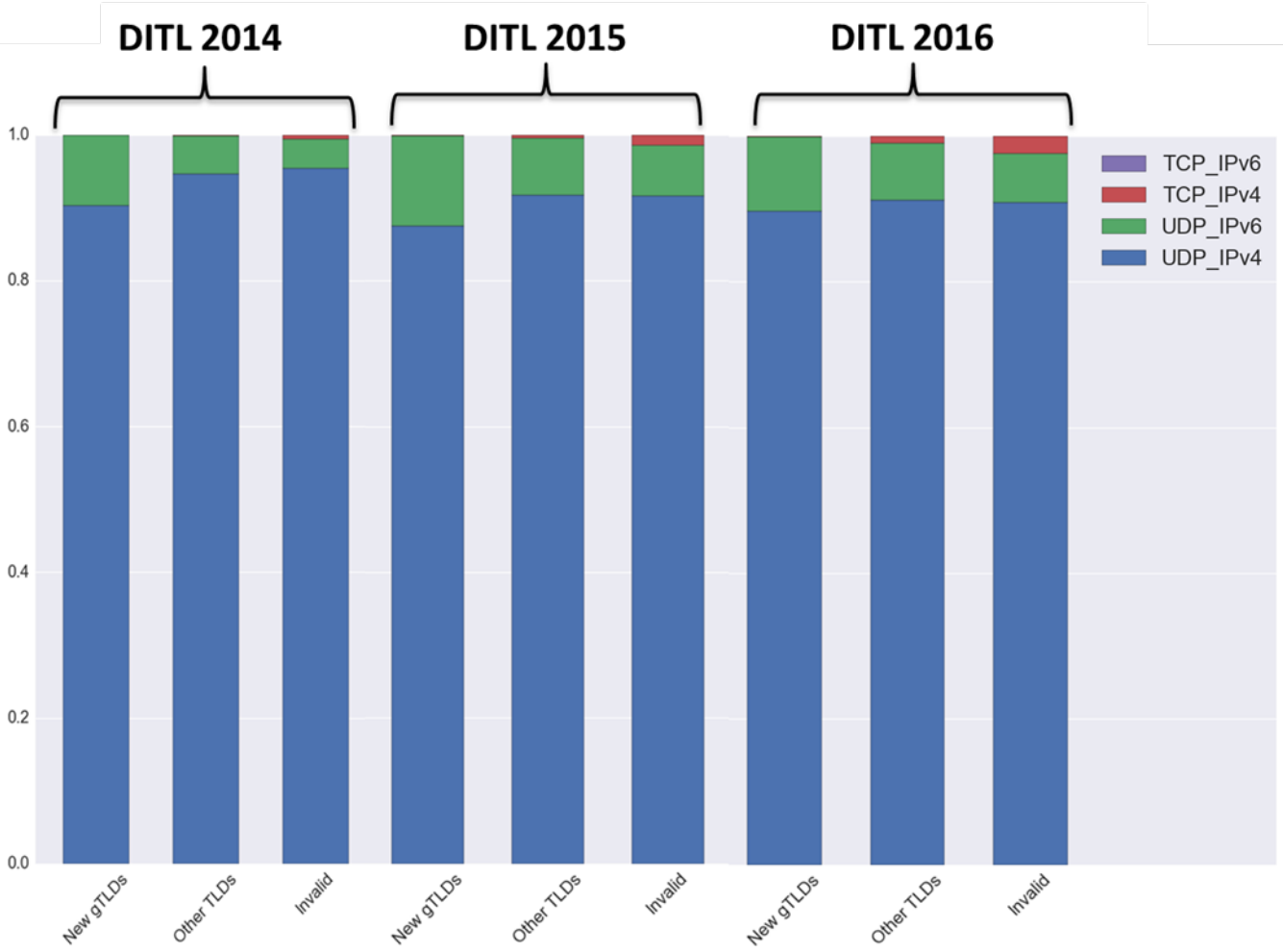
- DNSSEC broken chain validation



  - Structurally no more failures for new gTLDs than for other TLDs
  - Some single failures on startup for new gTLDs

  - https://meetings.icann.org/en/presentation-dnssec-monitoring-07mar16-en

Query type distribution (by transport)

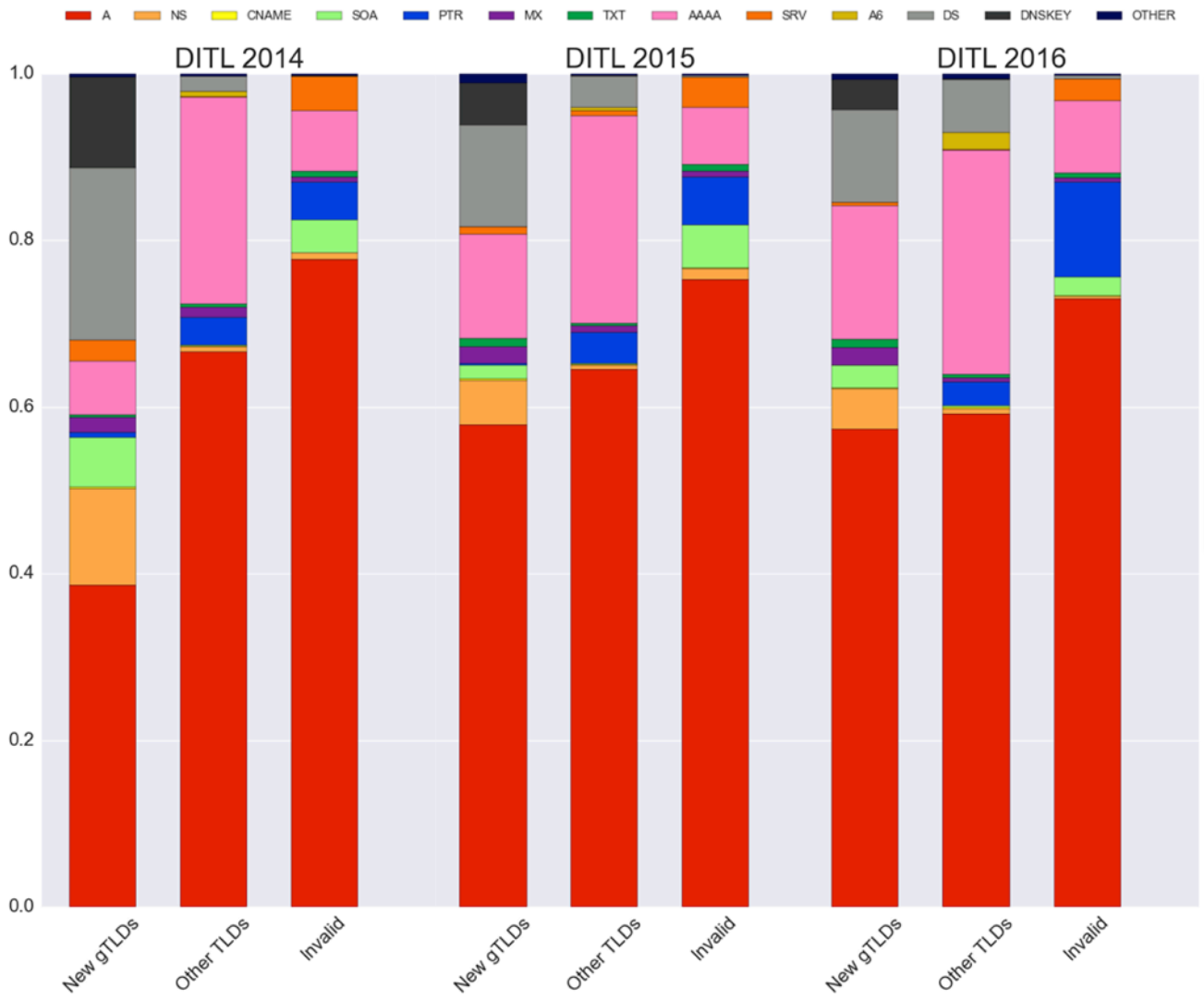


New gTLDs: queries to new gTLDs who are delegated at the time of the DITL set.

Other TLDs: queries to other TLDs who are delegated at the time of the DITL set.

Invalid: queries to names that have not been delegated.

Query type distribution (by RRType)



New gTLDs: queries to new gTLDs who are delegated at the time of the DITL set.

Other TLDs: queries to other TLDs who are delegated at the time of the DITL set.

Invalid: queries to names that have not been delegated.

NLnet Labs SIDN LABS TNO

A selection of geographic new gTLDs

Table 1 - Selection of geographic new gTLDs

TLD	Delegation date	Area type	Country
bayern	May 3rd 2014	Region	Germany
capetown	June 19th 2014	City	South Africa
doha	March 25th 2015	City	Qatar
london	March 22nd 2014	City	United Kingdom
melbourne	July 10th 2014	City	Australia
moscow	April 24th 2014	City	Russia
nyc	March 20th 2014	City	USA
rio	May 22nd 2014	City	Brasil
sydney	November 5th 2014	City	Australia
tirol	June 4th 2014	Region	Austria
tokyo	January 29th 2014	City	Japan
vlaanderen	June 18th 2014	Region	Belgium
xn--80adxhks	April 24th 2014	City	Russia

The DITL data sets differentiate into different anycast nodes, for a subset of Root Servers.

For some of these Root Servers, the anycast node names can be mapped to specific geographic locations. This allows us to use the DITL sets to derive query counts (split out per TLD) per location.

F-root, DITL 2016

A subset of F-root server anycast node locations

Within the anycast node in Frankfurt, the fraction of queries going to .bayern is **2.14 times** the fraction of **all** F-root queries going to .bayern.

	bayern	capetown	london	melbourne	sydney	moscow	xn--80adxhks	nyc	rio	tokyo
Frankfurt DE	2.14	1.04	2.08	0.752	1.28	0.533	0.449	0.971	1.13	0.488
Johannesburg ZA	0.706	3.92	0.74	0.357	0.246	0.0713	0	0.303	0.0305	0.181
London UK	1.09	3.59	3.9	2.65	2.84	1.29	1.59	1.25	3.43	1.01
Brisbane AU	0.941	2.06	1.42	9.56	11.5	0.638	0.867	1.48	2.16	0.722
Moscow RU	1.13	1.2	0.67	1.62	1.12	7.04	7.63	0.83	1.26	0.612
Atlanta US	3.66	1.68	7.03	1.63	1.11	0.742	0.881	1.76	0.893	1.01
Chicago US	1.15	1.11	2.53	0.93	0.849	0.818	0.704	2.33	0.687	0.637
Los Angeles US	2.92	3.55	1.46	4.81	6.4	2.79	2.14	2.13	4.59	1.14
New York US	1.3	1.29	0.746	1.83	1	0.982	0.499	2.44	1.08	0.969
Palo Alto US	1.26	1.64	1.07	1.26	1.31	0.664	0.537	1.43	1.06	1.11
San Jose US	1.13	1.07	2.36	2.18	1.52	1.32	0.412	3.77	0.395	1.05
SÃ£o Paulo BR	0.879	0.437	0.443	0.433	0.506	0.172	0.241	0.681	2.91	0.31
Osaka JP	0.887	0.395	0.298	0.409	0.524	0.204	0.214	0.628	0.836	16.7

The fraction of queries to a geographic new gTLD is higher than average in an anycast node whose location is in the TLD-related country (geographic affinity)

The highest increase is visible for .tokyo in the Osaka anycast node. But even there the fraction of queries going to .tokyo is microscopic: 0,015%

Observations about data

- DITL data
  - 'Umwelt' changes faster then current sampling
  - Data gets lost
  - History gets lost
    - Geographic info gets quickly unreliable

SAC046

Recommendation (4): ICANN should update its "Plan for Enhancing Internet Security, Stability, and Resiliency," to include actual measurement, monitoring, and datasharing capability of root zone performance, in cooperation with RSSAC and other root zone management participants to define the specific measurements, monitoring, and data sharing framework

DC: inability to obtain sufficient information to perform the modeling

KC & Vixie: inability to obtain sufficient information to perform the modeling

Suggested Next Steps

- Challenges for the DNS community:
  - Continuous improvement & standardization of data collection
    - Sanity checking of data is hard
  - More continuous measurement
  - More relevant detailed data
    - Long term H-root data was an "accident"
  - Better comparable public data
    - Noticed "phase errors" in 2014 DITL data
  - Better sample frequency

Summary

- Data quality varies and completeness can be improved
  - Encouraging is that standardization of data collection (DITL, RSSAC002) is improving
  - Collecting & analyzing per-TLD data provides new insights

- Preliminary conclusion
  - So far, we did not observe significant stability or security impact of new gTLD on RSS scale
  - NewGTLD traffic dwarfed by HOME/CORP/LOCAL traffic
  - In microscopic view some impact of new gTLDs is observed
    - query rate fluctuations / DNSSEC validation errors around initial delegation
  - Some geographic affinity for geographic new gTLDs is observed, but fraction of traffic to such new gTLDs remains insignificant

Questions and Discussion

CDAR Project Team
Bart Gijsen (TNO)
Benno Overeinder (NLnet Labs)
Cristian Hesselman (SIDN)
Daniël Worm (TNO)
Giovane Moura (SIDN)
Jaap Akkerhuis (NLnet Labs)

Coordinator
Bart Gijsen (Msc.)
+31 6 53 72 52 18
bart.gijsen@tno.nl

CDAR Home: http://www.cdar.nl