VERISIGN™

Regional Affinity for Applied for gTLD Strings

October 5, 2013

Presenter:
Andrew Simpson

Co-authors:
Danny McPherson
Eric Osterweil
Matt Thomas
Duane Wessels

What Are We Talking About?

- DNS Queries for some applied for strings originate disproportionately from certain countries
  - Root servers are currently responding to queries for the more than 1,400 applied for strings with an NXDomain response
  - This presentation outlines how more than 3 billion root queries were analyzed to determine regional affinity for specific applied for strings
- Our conclusion is that certain applied for strings are requested disproportionately by resolvers in some countries
  - By identifying the specific countries that have affinity for an applied for string, it is easier to further investigate what is generating these queries for the purpose of risk analysis

Overview

- Background
- Data Collection
- Regional Affinity Calculations
- Results

Background

- ICANN proposed new gTLD program
  - 1400+ strings have been applied for
- Security and Stability considerations with regards to delegation
  - What is the impact of delegating a new string in the root?
- Today's focus
  - Where are the users who are most likely to be impacted by each of the candidate delegations

Data Collection

- A and J root
  - a.root-servers.net. and j.root-servers.net.
  - Globally distributed resolution architecture
    - 17 primary resolution sites were instrumented
    - Remaining regional resolution sites are not yet included

- Observed window 7/16/2013 – 9/7/2013
  - The Verisign team has analyzed the most up-to-date data possible
  - Affinity based calculations rely heavily on establishing “normal” query patterns, and using this large dataset makes these baselines more reliable

Data Collected

- Instrumented sites pass root DNS responses through a span port
  - A process consumes these packets and extracts target features (currently filters down to queries about applied for strings only)
  - Features: FQDN, Timestamp, Destination IP (IP address where the query originated from)
  - Sample records:
    - 2013-07-24 04:37:16 v4IP#.#.#.# sld1.newtld1
    - 2013-07-24 04:37:16 v4IP#.#.#.# sld2.newtld1
    - 2013-07-24 04:37:16 v4IP#.#.#.# sld1.newtld1
    - 2013-07-24 04:37:35 v6IP#:#:#:#:#:#::# 4ld.3ld.sld3.newtld2
- What was collected
  - 3,811,657,217 queries analyzed
  - Queries for 1,409 applied for strings observed

Regional Data Assignment

- Destination IP Augmented with 2-letter country code using Maxmind GeoIP data
  - Aggregates are generated with raw query count by TLD by country

Applied for String	Country Code	Query Count
newtld1	AE	40
newtld1	AL	16
newtld1	AO	11
newtld1	AR	10
newtld1	AS	1
newtld2	AE	36
newtld2	AL	22
newtld2	AO	13
newtld2	AR	96
newtld2	AS	2

	Origin of Query				
Applied for String	AE	AL	AO	AR	AS
newtld1	40	16	11	10	1
newtld2	36	22	13	96	2
Region Totals	76	38	24	106	3

Normalizing for Regional Preferences

- On average, what proportion of the queries originating from a specific country are resolving a particular applied for string?

$$i_c^{AFS} = \frac{q_c^{AFS}}{Q_c}$$

$c = country$

$AFS = Applied\ for\ String$

$i_c^{AFS} = Proportion\ of\ queries\ for\ AFS\ from\ c$

$q_c^{AFS} = Number\ of\ queries\ for\ AFS\ from\ c$

$Q_c = Total\ queries\ from\ a\ c$

  - When Q_c is less than .01% of Q (the total observed query count) the queries from that country are not considered to avoid introducing volatility from countries where queries may not be well distributed

	Origin of Query (c)				
Applied for String	AE	AL	AO	AR	AS
newtld1	40	16	11	10	1
newtld2	36	22	13	96	2
Country Totals(Qc)	76	38	24	106	3

Applied for String	AE	AL	AO	AR	AS
newtld1	52.6%	42.1%	45.8%	9.4%	33.3%
newtld2	47.4%	57.9%	54.2%	90.6%	66.7%

  - Percentages serve to normalize query counts across countries

Establishing Baselines for Regional Preference

- The percentages serve as normalized values to compare countries for a given applied for string
  - The baseline for what is expected from a country is the average of all country proportions for an applied for string

$$I^{AFS} = \frac{\sum_{c=1}^{N} i_c^{AFS}}{N}$$

$I^{AFS} = Average\ of\ Country\ Percentages\ for\ an\ AFS$

$N = Number\ of\ Countries\ that\ meet\ minimum\ traffic\ threshold$

$i_c^{AFS} = Proportion\ of\ queries\ for\ AFS\ from\ a\ country$

  - The standard deviation of the proportions for an applied for string are then used to determine how far off the baseline any individual country is

% Distribution by TLD	Origin of Query (c)						
Applied for String	AE	AL	AO	AR	AS	Average	Standard Deviation
Newtld1	52.6%	42.1%	45.8%	9.4%	33.3%	36.7%	15.0%
newtld2	47.4%	57.9%	54.2%	90.6%	66.7%	63.3%	15.0%

Standard Deviations	Origin of Query (c)				
Applied for String	AE	AL	AO	AR	AS
newtld1	1.07	0.36	0.61	-1.82	-0.22
newtld2	-1.07	-0.36	-0.61	**1.82**	0.22

AR has an affinity for newtld2

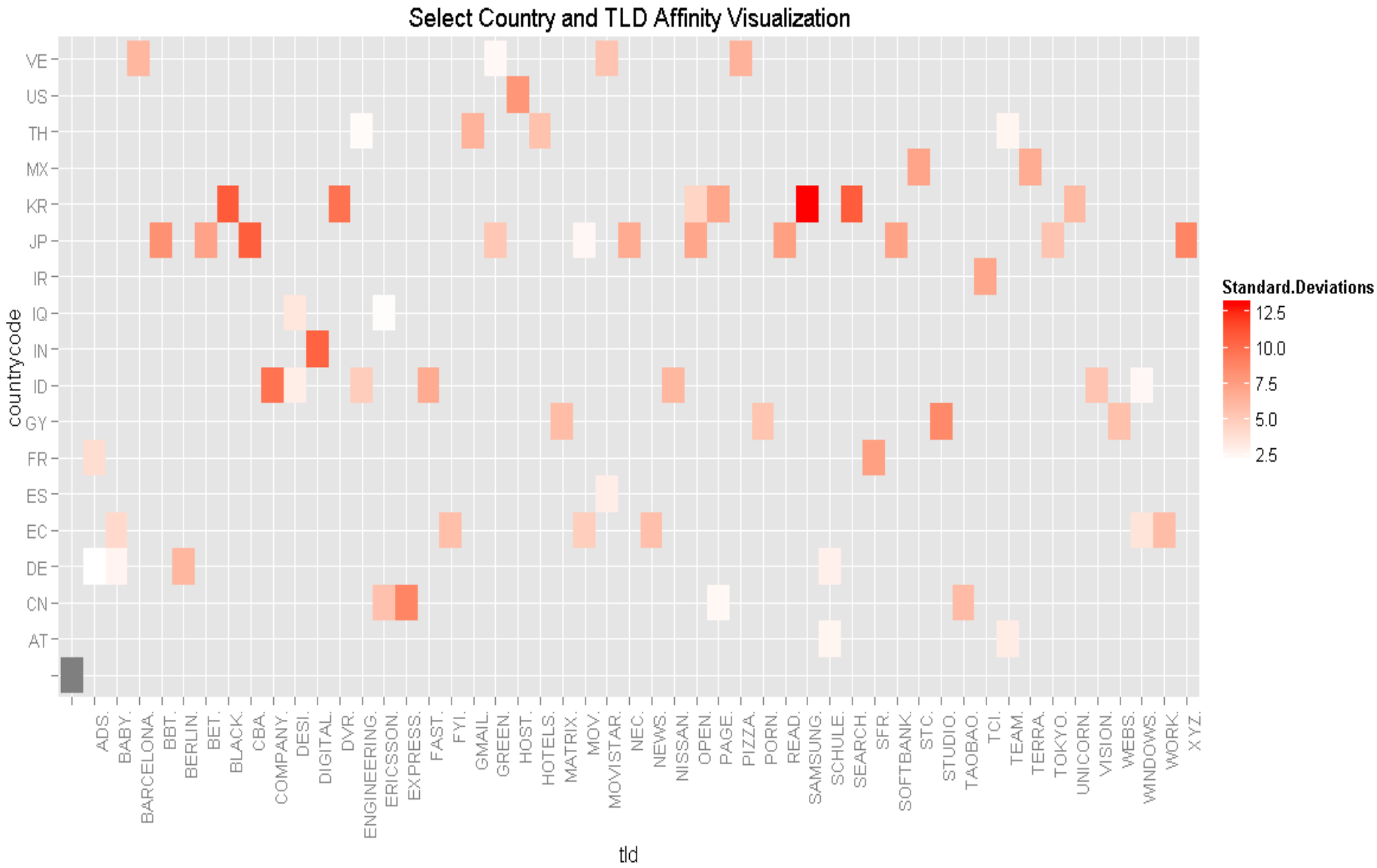
Raw Results

- Subset of full results

Originating Country/ Applied for String	Standard Deviations
DE	
.BERLIN	6.12
.SCHULE	2.86
.BABY	2.63
.COLOGNE	2.17
.HAUS	2.13
JP	
.CBA	10.69
.XYZ	8.85
.BBT	8.20
.READ	7.42
.BET	7.28
US	
.HOST	7.94
.WOW	5.17
.DENTAL	3.29
.COMCAST	2.75
.ANTHEM	2.37

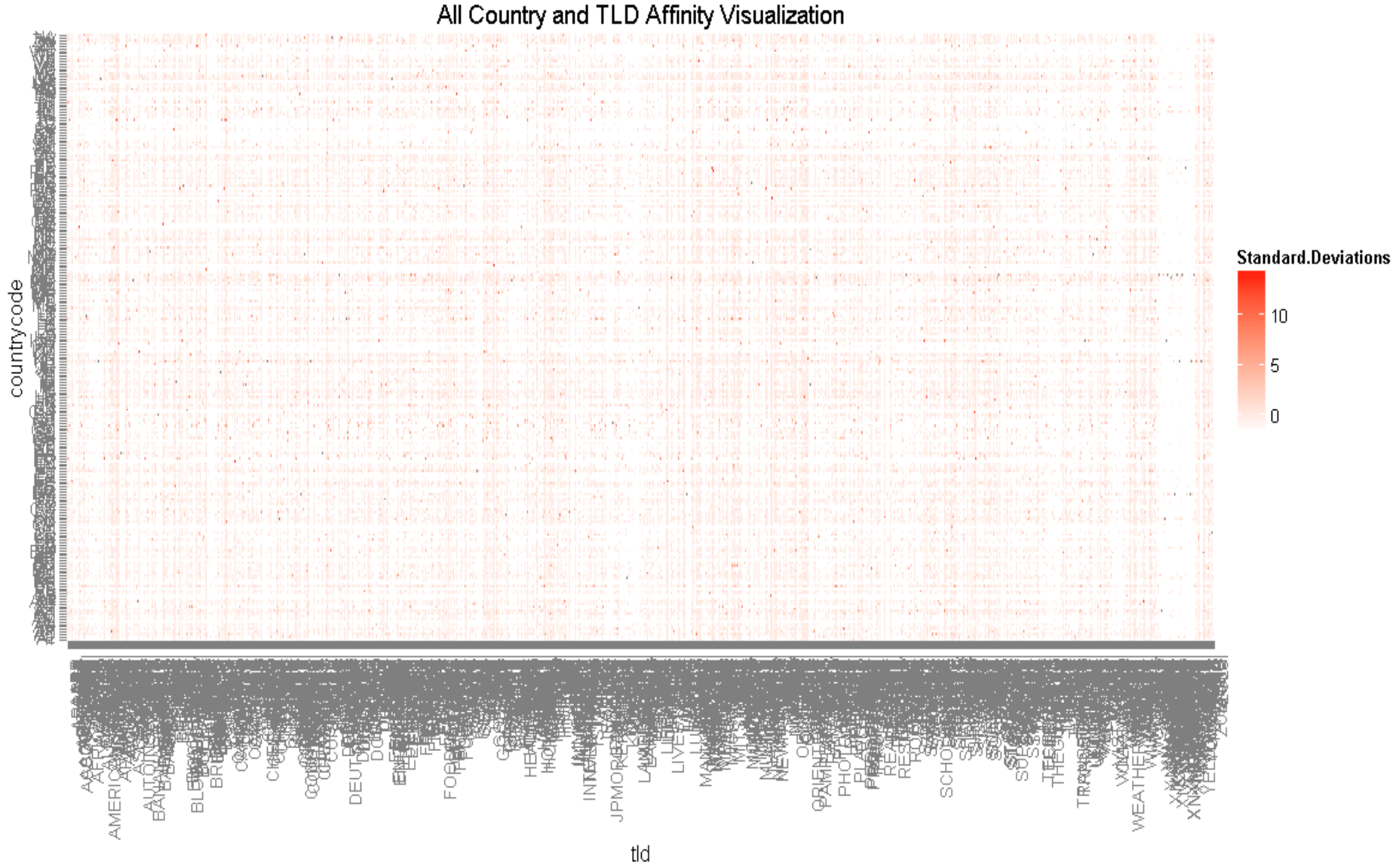
Originating Country/ Applied for String	Standard Deviations
FR	
.SFR	7.44
.BZH	5.05
.LOREAL	4.67
.ADS	3.98
.PROD	3.75
KR	
.SAMSUNG	13.04
.BLACK	10.81
.SEARCH	10.78
.DVR	9.77
.PAGE	7.10
ZA	
.MARRIOTT	4.35
.DURBAN	3.20
.EVENTS	3.19
.SKY	2.98
.CLOUD	2.36

Selected Results



Complete Results Visualization

http://www.verisignlabs.com/documents/Verisign%20Applied%20for%20String%20Regional%20Affinity.xlsx



Conclusion

- Determining risk associated with delegation of a new gTLD is not an easy exercise
  - Using this methodology it is possible to identify regions more likely to have more risk when delegating each new gTLD
  - If different regions are leveraging applied for strings differently, these results can help find the hotspots that need to be investigated
- Additional study
  - Better identification of affinity drivers
  - Repeat the same study with more data to analyze how the results continue to evolve

Thank You

VERISIGN™