Illustrating the Need to Undertake Qualitative Impact Assessments for Applied-For Strings: .WEBSITE, .COFFEE, and .CLUB

Our recent exploration of the query patterns for the .CBA applied-for string [1] have given us the opportunity to augment our broad based analysis of applied for gTLDs (from our *New gTLD Security, Stability, Resiliency Update: Exploratory Consumer Impact Analysis* [3]) with a focused methodology to identify potential impacted parties. Specifically, we composed a candidate methodology to relate detected *namespaces*, of various protocols with likely administrative boundaries. This enables us to automatically identify administrative boundaries even within and across Autonomous System (AS) boundaries or even routed prefix boundaries. In this study we examined 7 weeks of continuous queries from 12 global locations (100% of a root and 50% of j root server instances operated by Verisign). In this set we saw 11,841,909 queries for applied-for strings.

Initially, we used our methodology [1] to reveal previously undetected signals in the .CBA query stream. However, as a general methodology, we have begun using the same approach to examine query patterns seen for other applied-for strings. In particular, we have begun using our methodology (and the exact same data set used previously) to conduct qualitative study of applied-for strings that the Interisle report [2] has already classified; where two were classified as "*low risk*", and one was classified as "*uncalculated risk*." Our preliminary investigations into the strings .WEBSITE, .COFFEE, and .CLUB reveal measurable levels of previously undiagnosed risks.

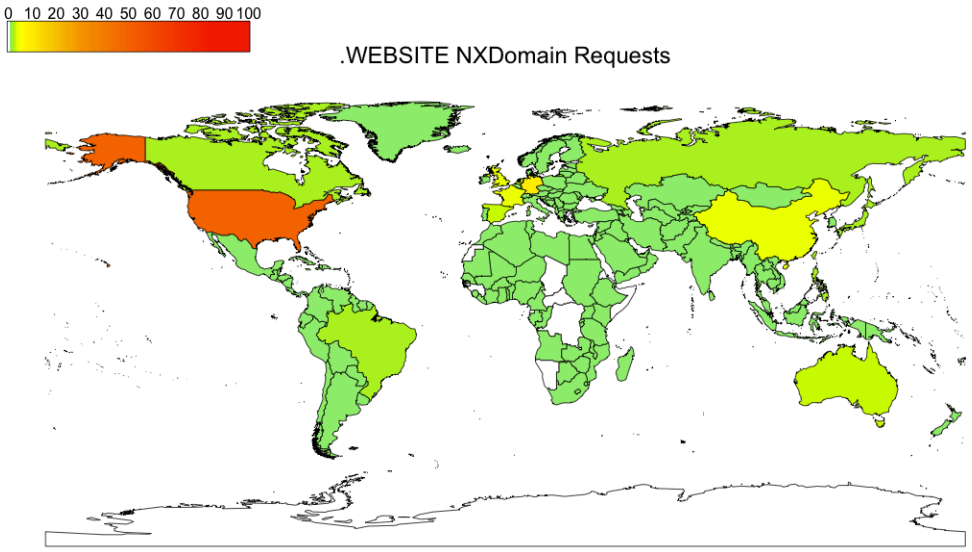

Figure 1 - This Figure shows a graphical representation of those countries that have been seen to issue queries for .WEBSITE. Of note is that any country with any color is potentially at risk for systemic effects of name-collisions for .WEBSITE.

The applied-for string .WEBSITE was ranked (by Interisle) as low risk, at #302. However, our dataset revealed that this applied-for string saw 411,200 queries from 205 countries [1] (as seen in

Figure 1), and from 5,918 discrete autonomous systems, or ASes. These numbers indicate a wide constituency for this applied-for string. Moreover, we saw 28,658 Second Level Domains (SLDs) in our data. While that number suggests diverse usage, our methodology also identifies namespaces that are used by actual protocols. Our analysis shows that 11 unique DNS Service Discovery (DNS-SD) namespaces are active under .WEBSITE, and they originate from 39 distinct ASes. In addition, the regions Faroe Islands, San Marino, Timor-Leste, and Samoa show statistically high regional affinities for the .WEBSITE applied-for string (our regional affinity scoring methodology is defined in [3]). Our investigation into specific reasons for these affinities is still ongoing, but the abnormally high preference from these regions suggests the possibility of systemic dependencies on this applied-for string.

Additionally, the Interisle study ranked the applied-for string .COFFEE even lower than .WEBSITE, Interisle (#627). This string was queried for 40,583 times from 111 countries (as seen in Figure 2), and from across 771 different ASes. Again, this spread strongly indicates a diverse geographical client-base for .COFFEE, and its 3,522 SLDs suggest a diverse semantic deployment. However, our methodology specifically identified five semantic namespaces conducting DNS-SD from 53 different ASes. While smaller in spread than .WEBSITE, these are 53 different autonomous networks might potentially be multiplied by as many as the number of unique namespaces seen under each. For simplicity, we conservatively take the maximum of these numbers (rather than their product) as a lower bound on the number of authorities seen for each applied-for string (that is, at least 53 different authorities appear to be responsible for DNS-SD queries to .COFFEE). In addition, this string receives regional affinities from Ethiopia and Hong Kong, and X.509 PKI Internal Name Certificates were seen for it in the 2010 SSL-Observatory dataset. The specific implications of these are described in our prior work [3]).

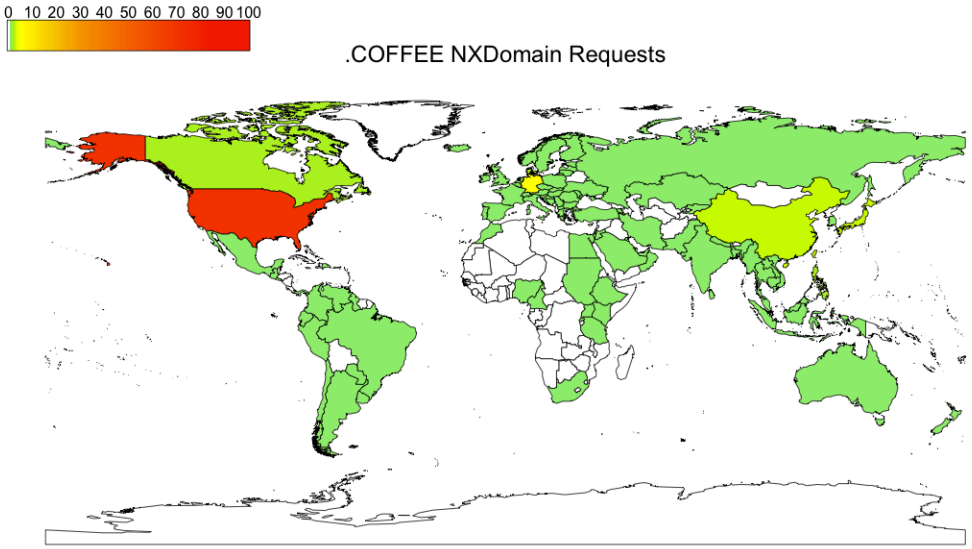

Figure 2 - This Figure shows a graphical representation of those countries that have been seen to issue queries for .COFFEE. Of note is that any country with any color is potentially at risk for systemic effects of name-collisions for .COFFEE.

Finally, we examined the applied-for string .CLUB, from the "*uncalculated risk*" category in the Interisle report. A recent comment, posted on ICANN's *Proposal to Mitigate Name Collision Risks* comments, [4] suggested that removing some of the SLDs with higher query volume would render this applied-for string less of a risk. However, this shallow assessment of risk falls into the same stunted categorization as the Interisle classification, or at least ICANN's rendering of their analysis. To illustrate the naïveté of this approach, we use qualitative analysis to illustrate the measurable dependencies on .CLUB. We observed 240,991 queries for this string (roughly five times the volume seen in the Interisle study, although with data from only ~15% of the root server system) from 173 countries (as illustrated in Figure 3) and we observed 21,157 unique SLDs under it. These measures of *spread* [3] suggest that a significant install-base is issuing queries. In addition, this string was observed to have 74 unique DNS-SD namespaces below it, and these were spread out over 146 ASes. Again, using the maximum of these values as a conservative lower bound (rather than their product), we can see that almost 100 unsuspecting impacted parties may be poised to becoming vulnerable to namespace collisions. In addition to the DNS-SD risks, our methodology also exposed that .CLUB demonstrates a more pronounced set of risks that either .WEBSITE or .COFFEE, as we observed queries for both WPAD and ISATAP, across 48 and 37 unique namespaces and from 116 and 173 ASes, respectively. These augment (rather than quantitatively add) to the risk vectors, as described in our prior work [3]. The presumption that examining roughly 48 hours of data, from essentially an arbitrary day, (i.e., the DITL data) in order to classify which SLDs are the most prominent is shockingly devoid of depth and quite a dangerous overgeneralization.

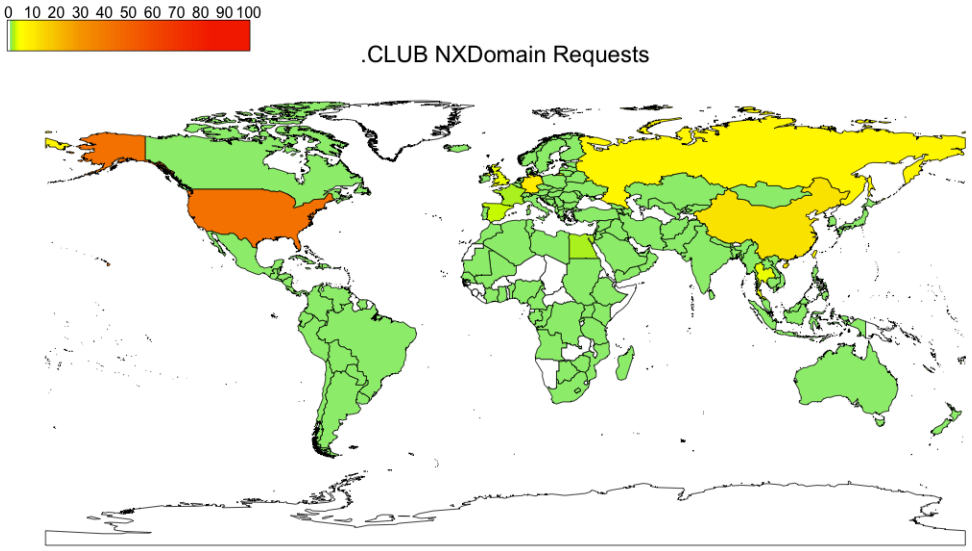

Figure 3 - This Figure shows a graphical representation of those countries that have been seen to issue queries for .CLUB. Of note is that any country with any color is potentially at risk for systemic effects of name-collisions for .CLUB.

In summary, two of these candidate strings are classified as *low risk* by the Interisle study. However, our transparent and public methodology [3, 1] details qualitative evidence that there is an active client-base spread across tens of (if not hundreds of) countries, which could potentially be impacted by delegation. This implies that the query-volume-based analysis being used as a classification scheme is fundamentally inadequate and strings must each receive individual study in order to classify the level of risk that is posed by delegation. Moreover, .CLUB (which was classified as

uncalculated risk by Interisle's initial ICANN-commissioned study) shows clear evidence of systemic usage from a wide swath of sources and countries. In a public comment, by .CLUB DOMAINS, LLC (and reportedly aided by Interisle), the Interisle methodology of evaluating query counts for just a fractional 48-hour window out of an entire year (i.e., the DITL data) is overextended to imply a direct correlation that prescribes the level of risk that would result from delegation of .CLUB. The initial results published by Interisle improperly suggest that there is a direct correlation between query counts and their qualitative meaning and impact. Furthermore, no consideration exists in that study to account for its unrepresentatively small sample-set of measurements. As a result, the reassessment of risk (by .CLUB DOMAINS, LLC) inherits the specious correlation between query volume and risk, first codified by ICANN's Proposal to Mitigate Name Collisions Risks. [5]

References

[1] Patrick Kane, "Focused Analysis on Applied-For gTLDs - .cba," http://forum.icann.org/lists/comments-name-collision-05aug13/msg00039.html

[2] Interisle Consulting Group, LLC, "Name Collision in the DNS," http://www.icann.org/en/about/staff/security/ssr/name-collision-02aug13-en.pdf

[3] Verisign Staff, "New gTLD Security, Stability, Resiliency Update: Exploratory Consumer Impact Analysis," Verisign Labs Technical Report #1130008, http://techreports.verisignlabs.com/tr-lookup.cgi?trid=1130008&rev=1

[4] .CLUB DOMAINS, LLC, ".CLUB Name Collision Resolution," forum.icann.org/lists/comments-name-collision-05aug13/msg00041.html

[5] "Proposal to Mitigate Name Collision Risks," http://www.icann.org/en/news/public-comment/name-collision-05aug13-en.htm