Neustar Insights | Report

A Methodology for Assessing Collision Risk and New gTLDs

neustar®
Real Intelligence. Better Decisions.™

CONTENTS

Executive Summary

Neustar views the preservation of stability and security of the domain name system ("DNS") to be the "prime directive" of the Internet Corporation for Assigned Names and Numbers ("ICANN") and all responsible participants in the DNS ecosystem. Accordingly, we have welcomed the community's focus on stability and security issues in connection with the launch of new generic top-level domains ("gTLDs").

In particular, we welcomed the report undertaken by Interisle Consulting Group, LLC entitled Name Collision in the DNS (the "Interisle Report"), which provides important insight into the possibility of a domain name collision. As the report itself acknowledges, however, query volume alone is not an adequate or appropriate basis for evaluating the risk of harm associated with any such collisions. As Interisle notes, the risk that ICANN must address and mitigate is the "potentially harmful consequences of name collision, and not the name collision itself."

ICANN's mitigation strategy rests entirely on the possibility of collision, not the consequences. As a result, ICANN's plan, in response to the Interisle Report, would relegate many demonstrably low-risk gTLDs to the nether world of "uncalculated risk" and impose further unwarranted delay in the launch of those gTLDs. ICANN's approach goes beyond simple prudence; it unnecessarily slows down the process of rolling out gTLDs, which enterprises have been working on for years. Prudence and due deliberation are always called for in a system upgrade of this magnitude, but ICANN's "uncalculated risk" category throws too many clearly low-risk gTLDs into a nightmare of uncertainty, and needs to be fixed.

Moreover, we disagree with the need for delay to conduct additional research in order to quantify the risk associated with the introduction of the vast majority of proposed new gTLDs. Rather, we believe that ICANN already has all the data and research necessary to calculate the risk and develop mitigation strategies that are carefully tailored to the specific risk associated with each TLD.

In this paper we propose an alternative, comprehensive risk evaluation methodology, based on an analysis of existing information available on four key variables including: (i) TLD query volume; (ii) query source IP address volume; (iii) queried second-level domain volume; and (iv) volume of SSL certificates.

Using these four inputs, one can calculate the relative risk for every applied-for TLD and compare that with known information about the many new TLDs launched without incident over the past decade. This analysis eliminates the "uncalculated risk" classification in the Interisle Report and the need for further research or qualitative analysis. Based on our analysis, Neustar identified only 3 TLDs that appear to merit mitigation strategies beyond the approach required for all other TLDs.

Finally, we offer a mitigation approach that reflects actual risk, is narrowly tailored to the type of risk involved, and in most cases eliminates the need for additional delay.

Introduction

The Interisle Consulting Group, LLC ("Interisle") studied and reported on the likelihood and potential consequences of collisions between new public gTLD labels and existing private uses of the same strings. The study, ***Name Collisions in the DNS***, was "concerned primarily with the measurement and analysis of the potential for name collision at the DNS root," rather than the risk associated with such collisions. As Interisle noted, however, it is the potentially harmful consequences of a collision and "not the name collision itself" that determines the risk arising from the introduction of any new gTLD string[1]. That was not Interisle's focus: rather, the "probability of the occurrence" was the study's "principle focus."[2]

The volume of queries for any particular string may be a reasonable measure of the ***possibility*** of ***collision***. As Interisle noted, however, the risk of harm associated with such potential collisions depends upon a variety of other factors including both additional data points and policy.[3] Based on its quantitative analysis of the incidence of name collisions, Interisle designated a small set of applied-for strings as "high risk." Similarly based exclusively on query volume, Interisle identified a number of strings as "low risk." Interisle concluded, however, that nearly twenty percent (20%) of the applied-for gTLD strings fell into the "uncalculated" category, presumably because Interisle concluded that using query volume as a proxy for risk of harm was inadequate in these cases.

ICANN's proposed mitigation plan, however, does just that—equating the volume of inquiries to the level of associated risk. Moreover, while both Interisle and the community have identified a range of mitigation strategies, ICANN's mitigation approach is based on an ***entirely different factor***—the amount of time requested by certificate authorities to address potential collisions with internal name certificates. While the Interisle study provides a valuable foundation for the conversation now underway in the community, it does not support ICANN's proposal to impose a 4-month hold on second-level registrations on all strings—including those for which certificates have not been issued—and its use of an arbitrary 80/20 rule to impose a further delay of unknown duration for several hundred strings.

Query volume alone is not an appropriate measure of risk. We believe, however, that existing data about the number of query sources, the appearance of strings as second-level domains (SLDs), and the existence of corresponding X.509 public key certificates can be combined with query volume data to undertake a more comprehensive evaluation of the relative "risk of harm" associated with collisions for virtually all of the applied-for strings.[4] A more nuanced risk assessment methodology also facilitates more focused and effective mitigation efforts and, in turn, the timely launch of new gTLDs.

From Query Volume to Risk Assessment

To calculate risk, one must understand both the magnitude of potential harm and the level of exposure to the chance of injury or loss.[5]

The research and analysis provided in the Interisle Report provides a solid foundation for understanding the potential for DNS collisions during the launch of new TLDs. It also reflects the community's commitment to ensuring that the launch of new gTLDs does not compromise the stability and security of the DNS. Notably, there have been no such studies in advance of introducing new TLDs in the last 12 years—and fortunately there have been no notable collisions with the launch of new TLDs during that time. As Paul Mockapetris, "father of the DNS" states, "There is an unprecedented level of caution."

Although the Interisle Report briefly describes some of the theoretical consequences of a name collision, it does not attempt to quantify such risk. Only 3 pages of the 178-page report address the potential consequences resulting from name collisions; they do not address the likelihood that such consequences would occur or the ramification of such collisions if they indeed did occur. The report outlines some of the tools that can be used to classify risk, but it does not implement them. It collects much of the data that it would need to measure impact, but then does not apply them to the risk formula. It proposes what the risk gradient chart would look like, but it does not actually complete it.

1) Interisle Report at 2-3.
2) Id. at 77.
3) Id. at 3.
4) To conduct this risk analysis, we leveraged the same OARC data made available to Interisle for its initial research, which has now been made available to new OARC members
5) http://dictionary.reference.com/browse/risk?s=t

That is not an oversight. The Interisle research had a very clear purpose: to identify the possibility of collision. Said another way, the report commissioned by ICANN was intended to examine whether it was theoretically possible that a domain name collision could occur in the new gTLD space, not to assess the impact of such a collision on security and stability. ICANN itself states:

The study was to consider whether name collisions might occur between applied-for new gTLD strings and non-delegated TLDs that may be in use in private namespaces. The study was also to review the possibility of name collisions arising from the use of X.509 digital certificates.[6]

The Interisle Report acknowledges this as well and explicitly notes that the possibility of a collision and the actual likelihood of risk of exposing the larger Internet to some danger are two separate and distinct items:

This study was concerned primarily with the measurement and analysis of the potential for name collision at the DNS root . . . the risk associated with delegating a new TLD label arises from the potentially harmful consequences of name collision, not the name collision itself.[7]

This paper takes the Interisle Report to the next level using additional variables that quantify the "severity of consequences" component of the risk equation and providing a holistic understanding of risk. We use many of the same elements that Interisle recommends as a proxy to understand the severity of consequences. We build upon the probability of collision research that is so thoroughly detailed and then supplement that research with a quantitative analysis of the consequences of a collision. The analysis provides an analytical methodology to both classify risk and then propose appropriate measures to mitigate any harmful consequences.

Query Volume is Not a Proxy for Risk of Harm

Several industry and thought leaders have noted that the exclusive use of query volume to non-existing domains (NXD) to assess risk of harm is flawed. Both Neustar and the New TLD Applicant Group (NTAG)—which includes the world's leading technology innovators as its members—pointed this out in the preliminary comment period.[8]

While Interisle might be right that fewer queries mean less risk, its assumption that a "reasonable threshold for low risk" could be established by reference to the number of queries for existing TLDs that are empty (meaning that their zones contain only the necessary DNS meta-data) is overly simplistic. Likewise, VeriSign confirms that "there is evidence that suggests that the traffic volume is not the only indicator of risk," when proposing their own risk profile model.[9]

Suffice it to say that domain name collisions happen every day. A typo in system code or a simple "fat finger" will often land the inquiring party, the one requesting the domain name, in a location that they were not expecting. These collisions have existed for years and will occur far into the future—whether or not there are new gTLDs. The resilience of the DNS in this respect is a relevant factor in evaluating the risk of harm from name collision: despite near constant name collisions, we are unaware of any such incident that threatened the security and stability of the Internet.[10]

Verisign, for example, has pointed to only two incidents of collision, neither of which affected the security and stability of the Internet, both of which occurred in the .COM top-level domain.[11]

The fact that collisions have occurred in .COM is not a surprise, given the number of .COM domain names and associated DNS traffic queries. In fact, the likelihood of a collision in the .COM name space is exponentially greater than in any one of the new gTLDs. Using data gathered from Neustar's UltraDNS Advantage platform, we compared queries for undefined name volumes for established and proposed TLDs during the 2013 DITL time period.

---

6) ICANN Mitigation Plan
7) Interisle Report at 2-3.
8) Commenters noted that ***Asia, .KP, .AX, .UM and .CW all saw higher query traffic than all 279 of the "Uncalculated Risk" strings; yet the launch of these strings proceeded without any known issues. These string traffic volumes are real world examples of new string delegation and provide a baseline: they used existing traffic, did not cause security and stability impacts, and can serve as a more effective threshold in classifying risk based on query volume***.
9) Verisign - New gTLD Security, Stability, Resiliency Update; Exploratory Consumer Impact Analysis at 3
10) Neustar notes that there was one documented issue within .xxx that involved a name collision (described at http://www.geek.com/news/just-launched-russian-itunes-full-of-porn-due-to-xxx-domain-snafu-1531240/), however, that issue most likely arose because a system designer decided to use an internal .xxx placeholder after the TLD was delegated and not before. No form of mitigation can prevent the collision of a name after the TLD has been delegated.
11) New gTLD Security, Stability, Resiliency Update: Exploratory Consumer Impact Analysis.

As the graph below illustrates, .COM experienced over 30,000 times more queries than did the string .NYC, 170,000 more inquiries than the string .SECURE, and 1.3 million times more queries than .CLUB. Despite the staggering difference in the volume of such queries, ICANN has relegated .NYC, .SECURE, and .CLUB to the category of "uncalculated risk" and put them on indefinite hold for further research and analysis. Meanwhile, .COM, with 442 million opportunities for domain name collisions, continues to register new domain names, respond to queries, and operate as normal.

TLD	Queries for Undefined Names	Multiple of .COM Query Volume
Com	442,804,764	
Home	74,508,207	6
Corp	3,186,780	139
Cisco	385,988	1,147
Inc	194,989	2,271
Office	122,880	3,604
Nyc	14,556	30,421
Dell	5,712	77,522
Secure	2,462	179,856
Club	337	1,313,961
Neustar	36	12,300,132

So why have we not heard the chorus of advocates proposing to cease new registration of .COM domain names while the community investigates risk? There two reasons:

1. Volume of queries for undefined names is not a good single measure of the risk resulting from a DNS collision; and

2. The Internet is resilient and able to support innovation through the introduction of new gTLDs.

Existing Data Supports Risk Profile Scoring

Both Interisle and ICANN call for putting the launch of hundreds of new gTLDs on hold pending further risk assessment and analysis. Based on our analysis, that is unnecessary.

Both the data provided by the DNS Operations Analysis and Research Center (OARC) as well as Interisle's Report can be used to evaluate risk and develop tailored mitigation strategies. This research has been supplemented by the community, which in recent weeks has provided significant funding for new OARC equipment and undertaken self-funded analysis. Data scientists, policy experts, and security analysts have invested countless hours to conduct an extensive analysis of this issue.[12]

A More Refined Risk Profile Methodology

Neustar is in the business of assessing risk. As the leading provider of information analytics, we help the largest retailers, ecommerce providers, financial institutions and government agencies identify, assess and mitigate risk every day. Understanding risk is what we do.

Accurately predicting risk is always challenging, but the fundamentals of risk assessment are simple. There are two key elements:

1. The likelihood of an event occurring

2. The impact of that event on the system

While the Interisle Report identified both elements, it only quantified the first. When the data is on hand, as it is here, it is possible to quantify the second element and develop and apply a methodology to derive a risk score.

12) Indeed, on the same day that the Interisle report was published, VeriSign published research that it described as "one of the largest investigations of DNS root zone traffic to date, with DNS queries from up to 11 of the 13 root instances, dating back to 2006."

Risk Score Inputs

A comprehensive Risk Assessment classification requires a compilation of risk factors that can assess, weigh and measure both the likelihood and impact of an event. Neither the Interisle Report nor the ICANN mitigation plan considered impact. Digicert, one of the world's leading certificate authorities, puts it best, stating that ICANN's "overly cautious approach is a result of purely considering the number of potential gTLD collisions without factoring in the other information provided in the Interisle report. We believe that when the additional data on certificates, SLD information, and total number of domains are considered, only a handful of strings truly need further consideration."[13]

Application of Neustar's risk profile methodology, as seen later, fully supports this belief. Our risk assessment module considers four distinct inputs:

1. NXD query volume (l_1)
2. Diversity of NXD request sources IP addresses (r_1)
3. Diversity of SLDs in NXD requests (r_2)
4. SSL certificates issued (r_3)

For each of the vectors above, we provide both the raw data and also a relative risk score associated with that data. Visualization of the data for these vectors supports Digicert's contention—only a handful of new gTLDs merit further analysis.

The following sections provide some detail on the 4 variables along with illustrations that help visualize the data. We also provide a risk scoring methodology in the summary findings.

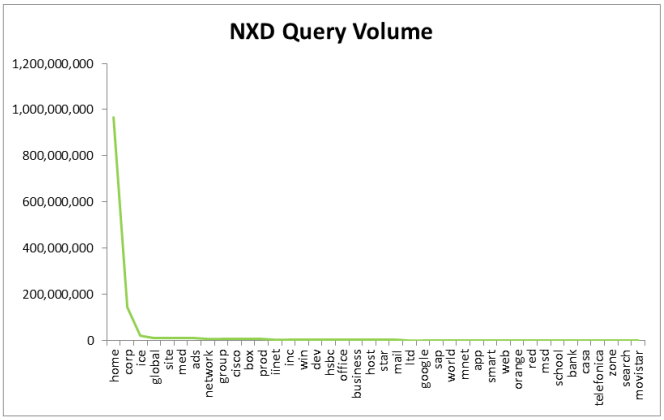
NXD Query Volume Score (l_1)

Neustar first used the relative NXD query volume of applied-for strings to calculate a Query Volume Score.

Rationale: NXD query volume, as the Interisle Report concluded, appears to provide an adequate proxy to determine the likelihood of the occurrence of a domain name collision. For purposes of the risk profile, the domain name collision is the risk event. Although DNS caching and other factors may obfuscate some of our ability to see every DNS query at the root, the NXD query volume provides the best estimation for the likelihood of collision.

Raw Data: To perform this analysis, Neustar and other OARC members reviewed data from the 2013 DITL dataset and produced accurate counts of query volumes. These counts were in line with the Interisle Report, with minor variations due to increased accuracy of tools employed.

The raw query counts reflect the dramatic difference between a very few TLDs and the remaining proposed TLDs. As discussed above, except for the top TLDs, these numbers compare favorably to previously launched TLDS[14]. Consistent with the Interisle Report, the top TLDs show a significant skewing of results toward .HOME and .CORP, highlighted by the graph to the right.



Risk Score Methodology: The NXD Query Volume Score is then calculated using existing query volumes at the root servers, as shown in the DITL 2013 dataset. To calculate the relative risk score for each TLD, we divided the TLD's query total during the sample period by the highest query volume of the proposed TLDs. The highest query volume proposed for TLD was .HOME, with a query count of 829K in 2013. The resulting value is multiplied by 100. The results from all of these scores are listed later in the Total Risk Score.

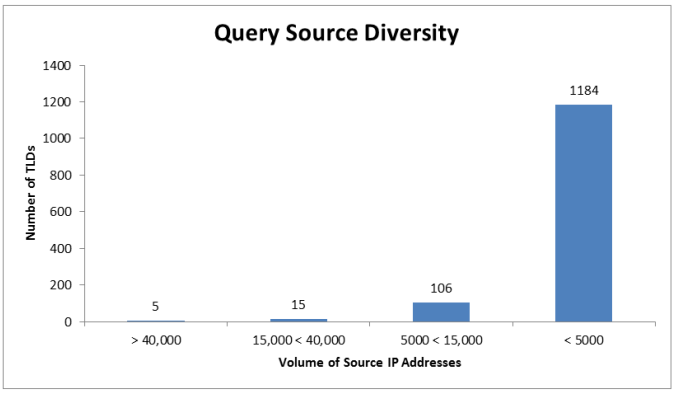
13) Jeremy Rowley, DigiCert Inc, DigiCert's Comments on new gTLD Collision Mitigation, August 27, 2013
14) Supra, Note 7.

Query Source Address Diversity Score (r_1)

Next, Neustar used information about source query diversity to calculate a Query Source Address Diversity Score.

Rationale: Query Source Address Diversity measures the number of different source addresses that are querying for a TLD. This provides the first of three indicators of the impact of the domain name collision. Measuring the relative number of source addresses and not the query volume of those source addresses eliminates a ratification inflation of the degree of risk driven by queries from the same source.

For reference, the number of source addresses for .SX, a TLD delegated in December 2011, exceeds 130,000. The number of source addresses for .SJ, the reference "cut-off" line for determining low or uncalculated risk in ICANN's mitigation recommendation, is over 5,500.



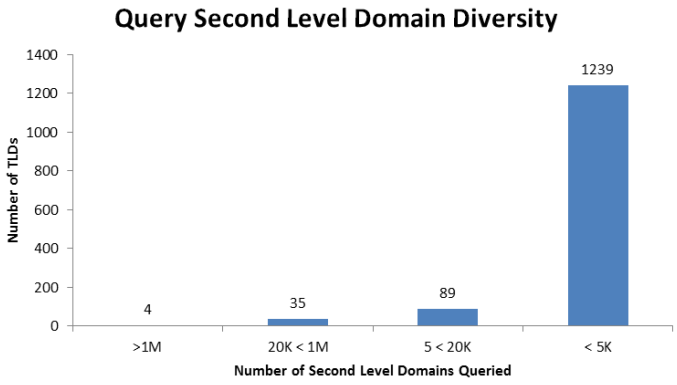
Raw Data: Analyzing the query source addresses, only a few TLDs have source address volumes of any significance. In fact, not one of the applied-for TLDs have more source addresses than the delegated ccTLD, .SX, and 90% of the (1257) have fewer source addresses than .SJ.

Risk Score Methodology: The Query Source Address Diversity Score was calculated using a comparison of unique source addresses in query volumes for a specific proposed TLD, as compared to the proposed TLD with the highest number of unique source addresses. The count of unique source addresses is calculated by identifying all source addresses for queries and their respective query counts. The top source addresses are then selected, accounting for a combined 98% of query volume. This approach was taken to identify primary querying resolvers. Ninety-eight percent was selected as a conservative value, to ensure most traffic was used in this count. To calculate the Query Source Address Diversity Score, this count is divided by the number of top 98% source addresses of the most diverse TLD, .MAIL with 66,006 unique source network addresses in the 2013 DITL data set, and the resulting value is multiplied by 100.

Query Second-Level Domain (SLD) Diversity Score (r_2)

Next, Neustar scored applied-for strings based on the relative number of second-level domains to which queries were directed.

Rationale: The Query SLD Diversity Score is an indicator of how many second-level domains within each new top-level domain are being queried. This data assists in quantifying the consequences of the collision. In some instances, a very few number of second-level domains account for the majority of the queries for that TLD. As an example, research for .NYC at the recursive DNS level revealed that 2 second-level domains accounted for 98% of the total queries for the TLD. Decoupling the number of second-level domain inquiries from queries at the root level provides a more accurate picture of the risk.



Raw Data: To support this analysis, Neustar and other OARC members reviewed data from the 2013 DITL dataset and for each TLD produced counts of unique SLDs queried. Once again, the raw data illustrates a sharp decline in risk associated with this vector after the first few names.

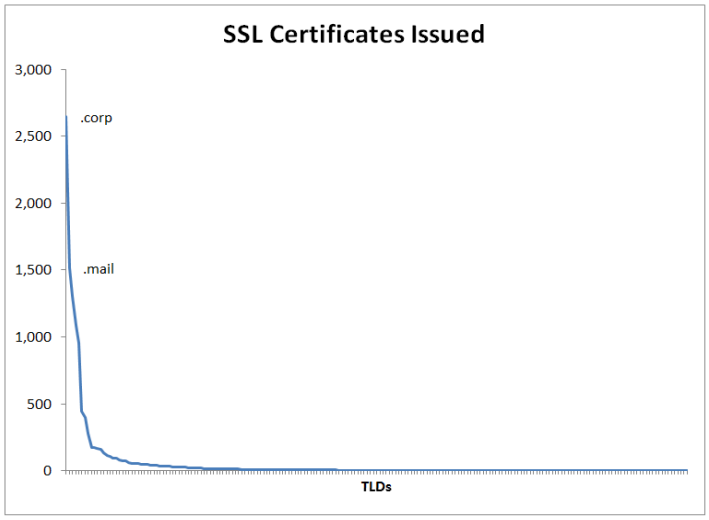
Risk Score Methodology: The Query SLD Diversity Score was calculated by comparing the unique second-level domains in query volumes for a specific proposed TLD to the proposed TLD with the highest number of these unique SLDs. The count of unique SLDs is calculated by first identifying all SLDs in queries and their respective query counts. Then the top SLDs are selected to account for a combined 98% of query volume. This approach was taken to identify SLDs queried with some level of frequency, while avoiding counting SLDs that may have been queried due to users mistyping network addresses. Ninety-eight percent was

selected as a conservative value, to ensure most traffic was used in this count. To calculate the Query SLD Diversity Score, this count is divided by the number of top 98% SLDs of the most diverse TLD in the 2013 DITL data set (.HOME with 441 million), and the resulting value is multiplied by 100.

SSL Certificates Issued Score (r_3)

Finally, Neustar scored applied-for strings based on the prevalence of corresponding SSL certificates.

Rationale: The final vector plays a significant role in understanding the consequences of a domain name collision. Many of the potential security issues involve the use of a digital certificate. ICANN even specified that the research of the digital certificates be included in the Interisle study. Digital certificates represent the ultimate illustration of end-user or system trust on the Internet. When end users or systems "see" a certificate, it provides assurance that they have in fact reached the place that they intended to reach. It also provides assurance that the user can provide information to that destination in a secure fashion.



The measure of the SSL Certificates Issued is exactly that, a representative count of the number of x.509 certificates issued by certificate authorities for proposed new TLDs.

Raw Data: The raw data for this analysis was gathered from the Interisle Report, Appendix C, offering a 2013 view of issued SSL certificates for proposed TLDs. For proposed TLDs that had fewer than 3 issued certificates, and thus were omitted from the Interisle report, a certificate count of 2 was used, as the most conservative value possible. The data again points to a relatively small number of TLDs accounting for the overwhelming majority of issued SSL certificates for new TLDs. In fact, 2 TLDs (.CORP and .MAIL) account for over 30% of the certificates issued corresponding to the 1300+ applied-for TLDs.

Risk Score Methodology: To calculate the SSL Certificate Issued Score for each TLD, the total of issued certificates corresponding to the TLD's string was divided by the highest total of issued certificates of the proposed TLDs—.CORP with a total of 2,747 certificates—and this resulting value was multiplied by 100.

Total Risk Score

Neustar then combined the risk vectors qualified above to create a numeric and normalized "Total Risk Score." We calculated the relative Total Risk Score for each TLD by multiplying the likelihood of the event occurring (l_1) by the consequences of the risk[15]. 7 In formulaic form:

$$\textit{Raw Risk Score} = (l_1) * ((r_1 + r_2) * r_3))$$

It is then helpful to put context to that risk score by normalizing the data to provide a comparative view of risk against the other TLDs in the data set. This is achieved by dividing by the maximum risk score and then multiplying by 100. In formulaic form:

Total Risk Score = Raw Risk Score / Max Risk Score *100

TLD	Queries	Query Score	# SSL Certs	# SSL Cert Score	# IPs (98%)	Source IP Score	# SLDs (98%)	# SLDs Score	Raw Risk Score	Risk Score
corp	138,562,192	15.523	2647	100.000	26987	40.886	5134	0.001	63,469.00	100.0000000
home	892,620,095	100.000	97	3.665	6007	9.101	441659734	100.000	39,980.23	62.9917472
mail	2,143,363	0.240	1517	57.310	66006	100.000	6791	0.002	1,376.16	2.1682321
ads	9,867,370	1.105	281	10.616	10952	16.592	744	0.000	194.72	0.3067891
global	11,171,889	1.252	169	6.385	12945	19.612	332	0.000	156.72	0.2469172
hsbc	3,331,164	0.373	1086	41.028	4471	6.774	10	0.000	103.71	0.1634049
dev	4,932,611	0.553	109	4.118	27704	41.972	14004	0.003	95.52	0.1504919
group	7,972,314	0.893	131	4.949	10390	15.741	394	0.000	69.58	0.1096246
inc	4,324,938	0.485	175	6.611	12250	18.559	1107	0.000	59.45	0.0936686
office	3,652,918	0.409	173	6.536	11575	17.536	100102	0.023	46.96	0.0739950
network	8,578,025	0.961	112	4.231	6524	9.884	1089	0.000	40.19	0.0633234
prod	6,884,212	0.771	51	1.927	13074	19.807	15268	0.003	29.44	0.0463813
site	8,238,487	0.923	32	1.209	10921	16.545	108171	0.024	18.49	0.0291297
star	2,170,524	0.243	39	1.473	7722	11.699	31	0.000	4.19	0.0066038
host	3,045,466	0.341	13	0.491	16307	24.705	309	0.000	4.14	0.0065224
exchange	146,366	0.016	1302	49.188	2341	3.547	11185	0.003	2.86	0.0045102
box	7,599,098	0.851	2	0.076	25934	39.290	311766	0.071	2.53	0.0039891
ltd	1,953,579	0.219	33	1.247	5675	8.598	347	0.000	2.35	0.0036961
app	1,294,141	0.145	8	0.302	33391	50.588	3138	0.001	2.22	0.0034925
web	1,059,832	0.119	22	0.831	13758	20.844	44401	0.010	2.06	0.0032423
bank	748,834	0.084	78	2.947	4784	7.248	576	0.000	1.79	0.0028230
tech	360,363	0.040	54	2.040	9800	14.847	908	0.000	1.22	0.0019266
cisco	7,347,752	0.823	2	0.076	12423	18.821	1833768	0.415	1.20	0.0018850
red	1,012,208	0.113	46	1.738	3997	6.056	35008	0.008	1.19	0.0018826
zone	691,689	0.077	24	0.907	10636	16.114	129	0.000	1.13	0.0017838
cba	115,309	0.013	952	35.965	1418	2.148	356	0.000	1.00	0.0015726
llc	587,266	0.066	47	1.776	5597	8.480	154	0.000	0.99	0.0015607
email	97,915	0.011	157	5.931	9598	14.541	4946	0.001	0.95	0.0014907
itau	144,242	0.016	442	16.698	2297	3.480	2468	0.001	0.94	0.0014797

15) A Word on Weighting. The model does not apply a qualitative weight to each of the risk factors. These weights can be adjusted based on qualitative assumptions (which would devalue the quantitative purpose of the analysis. By nature of the equation, two factors (Domain Name Queries and SSL Certificates) have the most impact on the outcome.

Appendix A lists the Total Risk Score for the first 150 TLDs. The top 30 TLDs are provided below.

What the Results Tell Us

There are two primary takeaways from this risk analysis:

1. The risk caused by the occurrence of a domain name collision can be calculated for every proposed TLD from the data that is available today, and
2. Only 3 TLDs stand out as having considerably higher risk than other proposed TLDs.

Risk Can Be Calculated

As mentioned earlier in this report, there is more data and analysis on the launch of new TLDs available to policy and decision makers than at any time in the history of the DNS. We have data from the Interisle Report, OARC data, and volumes of input from security and technical experts on domain name collision, enabling leaders to make one of the most informed decisions in the 12-year history of delegating top-level domains.

The risk model provided here provides a methodology for understanding and quantifying the relative risk across the new gTLD applicant pool. Each TLD can be provided a score that enables a comparative view of risk.

3 TLDs Merit Further Consideration

The data points to a dramatic difference between the risk score of the highest ranked TLDs and all other TLDs. The TLD with the highest risk is .CORP, with a risk score of 100. The second highest risk score is .HOME with a score of 62.99. The third highest TLD is .MAIL with a risk score of 2.17. Beyond these 3 TLDs, no single TLD has a score higher than 1.00, with most TLDS reflecting scores so insignificant that the data table needed to show 6 decimal places simply to illustrate that there was a quantifiable figure. One can reasonably assess that TLDs with a risk score lower than 1.00 can be grouped into a "low-risk" category for mitigation purposes. These TLDs represent a risk magnitude less than the top applied-for TLDs, .HOME, .CORP, and .MAIL.

Risk Classification and Mitigation Recommendations

The preceding research and resulting risk profile lends itself to a risk classification system that fits into the model proposed by ICANN. It further quantifies the risk for all proposed TLDs, placing them in one of two risk categories: High Risk Strings and Low Risk Strings.

- High Risk Strings: 3 TLDs (.HOME, .CORP, .MAIL)
- Low Risk Strings: All remaining TLDs

The risk calculation removes the need for an "uncalculated risk" category. The risk profile supports ICANN's initial finding of both the .HOME and .CORP TLDs as being high risk, but adds .MAIL into this category as well. All other TLDs on both the current low risk and uncalculated risk categories can be classified as low risk.

From this classification, registry operators, ICANN, and the larger Internet community can develop focused and tailored approaches to further reduce both the likelihood and, more importantly, the consequences arising from a domain name collision.

ICANN proposed a series of risk mitigation proposals and requested feedback from the community to help strengthen the effectiveness of those measures. In many cases, the mitigation approach described below provides a more surgical and targeted mitigation recommendation that addresses the specific risk as opposed to a broad sweeping measure that inhibits innovation and growth. The following section outlines ICANN's mitigation proposals along with Neustar's recommendations to help improve the effectiveness of those proposals.

High Risk Strings

- *ICANN Proposed Mitigation*. ICANN proposes to not delegate these strings until such time that the applicant can demonstrate that its proposed string should be classified as Low Risk.
- *Neustar Recommendation*. Neustar concurs with ICANN's approach. Based on the risk profile of these TLDs, it is prudent for industry leaders to take a more methodical approach to deploying these TLDs.

Low Risk Strings

1. *ICANN Proposed Mitigation: 120-day Wait Period*. "Registry operators will implement a period of no less than 120 days from the date that a registry agreement is signed before it may activate any names under the TLD in the DNS. This measure will help mitigate the risks related to the internal name certificates issue as described in the Study report and SSAC Advisory on Internal Name Certificates."

    - *Neustar Recommendations:*

    a. ***Commence certification revocation immediately.*** In collaboration with the CAB Forum, ICANN should work to begin the revocation of certificates for applied-for TLDs immediately. Waiting for contract signing unnecessarily increases the risk associated with potential collisions for reasons that are largely administrative. This would, in turn, provide even more time to help notify and fix systems that are utilizing the unverified domain name certificates without the risk of a domain collision occurring. The 120-day wait period would then commence upon notification of revocation from the CAB Forum or from contract signing, whichever is earlier.

    b. ***Apply the 120-day wait period (based on the recommendation above) only to those TLDs with a significant amount of certificates issued as identified by the Interisle report***. The 120-day wait period is intended to mitigate the risk associated with corresponding X.509 certificates. It provides a period of time for the revocation to take effect. For TLDs where no corresponding certificates have been issued, this 120-day period serves no purpose.

    c. ***Apply the domain name (SLD) activation restriction only to those names that account for the top 80% of NXD query volume***. The source of NXD query traffic can vary from a misconfigured system that is querying for the name in the DNS to a simple typo from an individual looking for a website. Those misconfigured systems generate volumes of NXD queries for SLDs. Conversely, the typos make up the long tail of the data, hundreds or thousands of queries for 1 domain name.

As a result, in many cases a small number of second-level domains represent an overwhelming majority of the NXD query traffic. Limiting the registrations of these names greatly reduces the probability of the domain name collision. The 80% provides a reasonable standard to determine query requests are from misconfigured systems and not typos.

2. *ICANN Proposed Mitigation: 30-day Notification Period*. "Once a TLD is first delegated within the public DNS root to name servers designated by the registry operator, the registry operator will not activate any names under the TLD in the DNS for a period of no less than 30 days. During this 30-day period, the registry operator will notify the point of contacts of the IP addresses that issue DNS requests for an un-delegated TLD or names under it."

    - *Neustar Recommendations:*

    a. ***Exclude second level registrations that allow registry operators to operate and promote its TLD from the 30-day hold.*** It makes little sense to withhold the delegation of all names within a TLD when the evidence does not suggest significant risk of collision. This would allow registry operators to use domain names for the operation and promotion of their TLDs as currently contemplated in the Registry Agreement, Specification 9 (Section 3.2).

    b. ***Remove the email notification requirement and replace it with a mandatory notification mechanism, such as a website, with information and instructions***. Collecting IP addresses for the purpose of notifying the administrators of those IP addresses has a host of challenges. First, the method is open to gaming in that malicious actors could generate queries for the sole purpose of generating work for the registry operator. Secondly, the administrative contacts listed for those IP addresses are often non-responsive or incorrect. Additionally, finding the actual end-user based on the source IP address is challenging given most corporate and ISP network architectures. Sending emails to those administrators would be ineffective in addressing the problem.

Alternatively, ICANN could mandate that for a period of time, new TLDs must post a standard educational website informing end-users that the TLD has been delegated along with information on how to update their systems to avoid future collision.

Conclusion

The introduction of new gTLDs has been delayed for years by an unending series of "what if" scenarios put forward by groups that never wanted new gTLDs in the first place. Since the new gTLD process began in earnest eight years ago, dozens of new gTLDs and ccTLDs have been launched, including a host of fast-track ccTLDs, .POST, .TEL, .ASIA, and .XXX. None of these launches were accompanied by harmful collision events, even though available data suggests that the potential for collision was relatively higher than it is for the new gTLDs. During that same period, prices for domain names have dramatically decreased, DNS providers have increasingly diversified, domains have become accessible to parts of the world that have never had access to domains before, and there has been increased innovation in the domain name market. Armed with the data we have, it's time to move forward.

Appendix A-Data Table

TLD	Queries	Query Score	# SSL Certs	# SSL Cert Score	# IPs (98%)	Source IP Score	# SLDs (98%)	# SLDs Score	Raw Risk Score	Risk Score
corp	138,562,192	15.523	2647	100.000	26987	40.886	5134	0.001	63,469.00	100.0000000
home	892,620,095	100.000	97	3.665	6007	9.101	441659734	100.000	39,980.23	62.9917472
mail	2,143,363	0.240	1517	57.310	66006	100.000	6791	0.002	1,376.16	2.1682321
ads	9,867,370	1.105	281	10.616	10952	16.592	744	0.000	194.72	0.3067891
global	11,171,889	1.252	169	6.385	12945	19.612	332	0.000	156.72	0.2469172
hsbc	3,331,164	0.373	1086	41.028	4471	6.774	10	0.000	103.71	0.1634049
dev	4,932,611	0.553	109	4.118	27704	41.972	14004	0.003	95.52	0.1504919
group	7,972,314	0.893	131	4.949	10390	15.741	394	0.000	69.58	0.1096246
inc	4,324,938	0.485	175	6.611	12250	18.559	1107	0.000	59.45	0.0936686
office	3,652,918	0.409	173	6.536	11575	17.536	100102	0.023	46.96	0.0739950
network	8,578,025	0.961	112	4.231	6524	9.884	1089	0.000	40.19	0.0633234
prod	6,884,212	0.771	51	1.927	13074	19.807	15268	0.003	29.44	0.0463813
site	8,238,487	0.923	32	1.209	10921	16.545	108171	0.024	18.49	0.0291297
star	2,170,524	0.243	39	1.473	7722	11.699	31	0.000	4.19	0.0066038
host	3,045,466	0.341	13	0.491	16307	24.705	309	0.000	4.14	0.0065224
exchange	146,366	0.016	1302	49.188	2341	3.547	11185	0.003	2.86	0.0045102
box	7,599,098	0.851	2	0.076	25934	39.290	311766	0.071	2.53	0.0039891
ltd	1,953,579	0.219	33	1.247	5675	8.598	347	0.000	2.35	0.0036961
app	1,294,141	0.145	8	0.302	33391	50.588	3138	0.001	2.22	0.0034925
web	1,059,832	0.119	22	0.831	13758	20.844	44401	0.010	2.06	0.0032423
bank	748,834	0.084	78	2.947	4784	7.248	576	0.000	1.79	0.0028230
tech	360,363	0.040	54	2.040	9800	14.847	908	0.000	1.22	0.0019266
cisco	7,347,752	0.823	2	0.076	12423	18.821	1833768	0.415	1.20	0.0018850
red	1,012,208	0.113	46	1.738	3997	6.056	35008	0.008	1.19	0.0018826
zone	691,689	0.077	24	0.907	10636	16.114	129	0.000	1.13	0.0017838
cba	115,309	0.013	952	35.965	1418	2.148	356	0.000	1.00	0.0015726
llc	587,266	0.066	47	1.776	5597	8.480	154	0.000	0.99	0.0015607
email	97,915	0.011	157	5.931	9598	14.541	4946	0.001	0.95	0.0014907
itau	144,242	0.016	442	16.698	2297	3.480	2468	0.001	0.94	0.0014797
google	1,095,466	0.123	2	0.076	62178	94.201	45446	0.010	0.87	0.0013764
cloud	485,121	0.054	48	1.813	5730	8.681	1357	0.000	0.86	0.0013480
sbs	164,450	0.018	396	14.960	1977	2.995	3949	0.001	0.83	0.0013011
win	5,219,695	0.585	7	0.264	2699	4.089	26	0.000	0.63	0.0009963
school	830,291	0.093	29	1.096	3536	5.357	68422	0.015	0.55	0.0008626
media	164,210	0.018	59	2.229	7942	12.032	14087	0.003	0.49	0.0007776
youtube	557,548	0.062	2	0.076	48782	73.905	23	0.000	0.35	0.0005495
world	1,561,287	0.175	4	0.151	8435	12.779	7356	0.002	0.34	0.0005323
law	284,858	0.032	40	1.511	4609	6.983	1866	0.000	0.34	0.0005306
you	517,307	0.058	2	0.076	40215	60.926	4452	0.001	0.27	0.0004203
city	289,684	0.032	27	1.020	4992	7.563	4861	0.001	0.25	0.0003945
sap	1,746,329	0.196	2	0.076	10975	16.627	6	0.000	0.25	0.0003873
med	593,439	0.066	6	0.227	10542	15.971	713	0.000	0.24	0.0003792
college	248,215	0.028	21	0.793	6747	10.222	775	0.000	0.23	0.0003553
live	350,155	0.039	6	0.227	15973	24.199	5134	0.001	0.22	0.0003390
services	224,960	0.025	42	1.587	3478	5.269	383	0.000	0.21	0.0003320
one	462,282	0.052	9	0.340	7796	11.811	2627	0.001	0.21	0.0003277
data	358,940	0.040	16	0.604	5276	7.993	5176	0.001	0.19	0.0003062
goo	294,527	0.033	2	0.076	43132	65.346	487	0.000	0.16	0.0002567
company	252,221	0.028	12	0.453	7753	11.746	1974	0.000	0.15	0.0002371

top	460,539	0.052	8	0.302	5973	9.049	560	0.000	0.14	0.0002223
abc	419,396	0.047	5	0.189	10076	15.265	26027	0.006	0.14	0.0002135
mnet	1,499,473	0.168	2	0.076	7022	10.638	14	0.000	0.14	0.0002127
telefonica	420,413	0.047	91	3.438	492	0.745	26841	0.006	0.12	0.0001917
cam	351,244	0.039	3	0.113	17779	26.935	13279	0.003	0.12	0.0001893
yahoo	342,659	0.038	2	0.076	26936	40.808	2996	0.001	0.12	0.0001865
blog	421,393	0.047	3	0.113	13202	20.001	72681	0.016	0.11	0.0001687
link	326,415	0.037	2	0.076	23121	35.029	13355	0.003	0.10	0.0001525
mobile	145,019	0.016	14	0.529	7004	10.611	7951	0.002	0.09	0.0001437
family	175,523	0.020	21	0.793	3827	5.798	18532	0.004	0.09	0.0001426
bet	476,086	0.053	2	0.076	13635	20.657	4185	0.001	0.08	0.0001312
new	478,134	0.054	2	0.076	13164	19.944	6515	0.001	0.08	0.0001272
hosting	148,398	0.017	15	0.567	5381	8.152	406	0.000	0.08	0.0001210
off	415,516	0.047	4	0.151	7171	10.864	3016	0.001	0.08	0.0001204
ecom	486,312	0.054	2	0.076	12155	18.415	320	0.000	0.08	0.0001194
farm	145,652	0.016	30	1.133	2695	4.083	1574	0.000	0.08	0.0001190
gmail	242,744	0.027	2	0.076	22085	33.459	3631	0.001	0.07	0.0001083
orange	988,770	0.111	9	0.340	1149	1.741	362544	0.082	0.07	0.0001082
secure	62,014	0.007	54	2.040	3193	4.837	4682	0.001	0.07	0.0001080
hermes	76,362	0.009	77	2.909	1760	2.666	4242	0.001	0.07	0.0001046
goog	196,853	0.022	2	0.076	25836	39.142	380	0.000	0.07	0.0001028
free	460,664	0.052	2	0.076	10544	15.974	14889	0.003	0.06	0.0000982
hot	536,438	0.060	2	0.076	8440	12.787	1029	0.000	0.06	0.0000915
life	43,350	0.005	32	1.209	6244	9.460	2844	0.001	0.06	0.0000875
here	223,427	0.025	2	0.076	18280	27.694	935	0.000	0.05	0.0000825
gold	339,747	0.038	15	0.567	1387	2.101	140218	0.032	0.05	0.0000725
work	347,194	0.039	3	0.113	6180	9.363	37742	0.009	0.04	0.0000651
show	238,248	0.027	2	0.076	13154	19.928	519	0.000	0.04	0.0000633
apple	230,183	0.026	2	0.076	13368	20.253	52655	0.012	0.04	0.0000622
amazon	175,910	0.020	3	0.113	11240	17.029	411	0.000	0.04	0.0000599
msd	955,439	0.107	2	0.076	3051	4.622	13	0.000	0.04	0.0000589
anz	72,299	0.008	31	1.171	2601	3.941	33	0.000	0.04	0.0000589
earth	169,793	0.019	15	0.567	2231	3.380	663	0.000	0.04	0.0000574
matrix	179,205	0.020	11	0.416	2831	4.289	30823	0.007	0.04	0.0000565
llp	131,108	0.015	30	1.133	1345	2.038	2399	0.001	0.03	0.0000535
store	91,560	0.010	10	0.378	5735	8.689	3372	0.001	0.03	0.0000531
center	307,736	0.034	6	0.227	2760	4.181	3372	0.001	0.03	0.0000515
hotel	454,190	0.051	2	0.076	5217	7.904	67733	0.015	0.03	0.0000480
zip	182,462	0.020	2	0.076	12770	19.347	33187	0.008	0.03	0.0000471
online	158,820	0.018	3	0.113	9528	14.435	12295	0.003	0.03	0.0000459
plus	300,177	0.034	2	0.076	7251	10.985	1500	0.000	0.03	0.0000440
bom	142,073	0.016	2	0.076	15310	23.195	6743	0.002	0.03	0.0000440
wiki	53,077	0.006	9	0.340	8555	12.961	4599	0.001	0.03	0.0000413
hotmail	134,751	0.015	2	0.076	14717	22.296	3498	0.001	0.03	0.0000401
art	320,154	0.036	2	0.076	5987	9.070	6284	0.001	0.02	0.0000387
wow	392,730	0.044	8	0.302	1181	1.789	209146	0.047	0.02	0.0000385
green	39,265	0.004	27	1.020	3504	5.309	6507	0.001	0.02	0.0000375
jinet	4,843,420	0.543	2	0.076	88	0.133	1901269	0.430	0.02	0.0000364
aaa	163,211	0.018	2	0.076	10824	16.399	13144	0.003	0.02	0.0000357
support	64,166	0.007	14	0.529	3709	5.619	6009	0.001	0.02	0.0000337

ice	20,155,369	2.258	3	0.113	54	0.082	2	0.000	0.02	0.0000330
shop	140,004	0.016	2	0.076	11469	17.376	6189	0.001	0.02	0.0000324
business	2,728,608	0.306	4	0.151	133	0.201	1054077	0.239	0.02	0.0000320
casa	640,314	0.072	2	0.076	2416	3.660	182059	0.041	0.02	0.0000316
aol	143,830	0.016	2	0.076	10412	15.774	6782	0.002	0.02	0.0000303
nyc	205,295	0.023	4	0.151	3598	5.451	1594	0.000	0.02	0.0000299
delta	163,012	0.018	7	0.264	2512	3.806	57631	0.013	0.02	0.0000291
olympus	88,229	0.010	13	0.491	2498	3.785	6021	0.001	0.02	0.0000290
pub	232,180	0.026	2	0.076	6127	9.282	418	0.000	0.02	0.0000287
auto	172,707	0.019	2	0.076	8016	12.144	3453	0.001	0.02	0.0000280
mit	112,001	0.013	4	0.151	6160	9.332	4026	0.001	0.02	0.0000279
bing	90,317	0.010	2	0.076	15063	22.821	121	0.000	0.02	0.0000275
vet	241,597	0.027	2	0.076	5500	8.333	305	0.000	0.02	0.0000268
page	208,290	0.023	2	0.076	6097	9.237	16088	0.004	0.02	0.0000257
csc	221,654	0.025	2	0.076	5697	8.631	254	0.000	0.02	0.0000255
news	131,333	0.015	2	0.076	9497	14.388	5538	0.001	0.02	0.0000252
car	88,308	0.010	3	0.113	9130	13.832	5072	0.001	0.02	0.0000244
sina	158,603	0.018	2	0.076	7430	11.257	1864	0.000	0.02	0.0000238
comcast	317,998	0.036	2	0.076	3615	5.477	41146	0.009	0.01	0.0000233
now	102,340	0.011	2	0.076	10792	16.350	3759	0.001	0.01	0.0000223
ski	84,255	0.009	2	0.076	12670	19.195	729	0.000	0.01	0.0000216
samsung	416,836	0.047	2	0.076	2534	3.839	76519	0.017	0.01	0.0000214
cal	75,696	0.008	4	0.151	6906	10.463	2825	0.001	0.01	0.0000211
bar	150,678	0.017	2	0.076	6889	10.437	2270	0.001	0.01	0.0000210
medical	365,950	0.041	2	0.076	2752	4.169	385	0.000	0.01	0.0000203
svr	108,895	0.012	8	0.302	2285	3.462	287	0.000	0.01	0.0000201
navy	126,361	0.014	2	0.076	7301	11.061	358	0.000	0.01	0.0000186
xyz	150,559	0.017	2	0.076	6080	9.211	2138	0.000	0.01	0.0000185
dell	182,273	0.020	2	0.076	4981	7.546	24560	0.006	0.01	0.0000184
lol	97,712	0.011	2	0.076	9260	14.029	9669	0.002	0.01	0.0000183
house	197,988	0.022	2	0.076	4502	6.821	22454	0.005	0.01	0.0000180
storage	19,943	0.002	76	2.871	1174	1.779	3690	0.001	0.01	0.0000180
man	161,047	0.018	2	0.076	5516	8.357	5245	0.001	0.01	0.0000180
search	623,960	0.070	2	0.076	1385	2.098	2350	0.001	0.01	0.0000175
london	79,174	0.009	15	0.567	1453	2.201	2123	0.000	0.01	0.0000174
foo	527,270	0.059	2	0.076	1512	2.291	87	0.000	0.01	0.0000161
lanxess	297,617	0.033	2	0.076	2593	3.928	3	0.000	0.01	0.0000156
srt	92,792	0.010	2	0.076	8231	12.470	1997	0.000	0.01	0.0000154
nexus	45,597	0.005	12	0.453	2722	4.124	4619	0.001	0.01	0.0000151
design	115,873	0.013	2	0.076	6404	9.702	5506	0.001	0.01	0.0000150
baidu	103,661	0.012	2	0.076	6877	10.419	358	0.000	0.01	0.0000144
fox	113,156	0.013	2	0.076	6152	9.320	2899	0.001	0.01	0.0000141
and	76,749	0.009	2	0.076	8588	13.011	9346	0.002	0.01	0.0000133
run	80,430	0.009	2	0.076	8092	12.259	944	0.000	0.01	0.0000132
thai	262,563	0.029	2	0.076	2442	3.700	40	0.000	0.01	0.0000130
computer	134,906	0.015	2	0.076	4739	7.180	20505	0.005	0.01	0.0000129
ibm	175,852	0.020	2	0.076	3500	5.303	5017	0.001	0.01	0.0000124
acer	154,092	0.017	2	0.076	3962	6.002	28912	0.007	0.01	0.0000123
team	98,098	0.011	3	0.113	4052	6.139	3368	0.001	0.01	0.0000120
sex	73,114	0.008	2	0.076	7932	12.017	8311	0.002	0.01	0.0000117
taobao	270,721	0.030	2	0.076	2033	3.080	56611	0.013	0.01	0.0000112

FOR MORE INFORMATION

Visit **www.neustar.biz**

About Neustar

Neustar, Inc. (NYSE: NSR) is a trusted, neutral provider of real-time information and analysis to the communications services, financial services, retail, media and advertising sectors. Neustar applies its advanced, secure technologies to help its clients promote and protect their businesses. More information is available at **www.neustar.biz.**

21575 Ridgetop Circle, Sterling, VA 20166
+1 571.434.5400 / www.neustar.biz

neustar®
Real **Intelligence.** Better **Decisions.**™