17 September 2013

Dear ICANN Board and Staff,

We, the members of the New TLD Applicant Group (“NTAG”), write to expand upon our initial comment about name collision and Security, Stability and Resiliency (“SSR”) and to reply to some comments we disagree with that were submitted during the initial 21-day commenting period for this topic. This letter reflects the interests of the entire NTAG. It is the result of an unprecedented, cooperative effort between rival applicants and with organizations outside of the ICANN community that are concerned about cyber security.

But before diving into the heart of this letter, we invite you to take a step back. Stability and security with the new gTLD program are core to ICANN’s mission and vital to the success of the entire program. If any gTLD launch goes awry — especially if due to SSR issues — it could spoil consumer adoption for all applicants who’ve launched or are waiting to launch. This is why NTAG believes it is in our interest to support enhancements to cyber security, including to convey data-driven analysis that helps you to decide what is best for the new gTLD program from a SSR perspective.

Since submitting the first letter, NTAG has been hard at work to provide you with what we hope is a substantial, data-driven update and reasonable mitigation approach:

- NTAG members have donated thousands of dollars worth of hardware to help OARC, the provider of root query data, to upgrade its capabilities to allow for greater access to and further study of Day in The Life of the Internet (“DITL”) data that is the subject of the Interisle report.
- NTAG has extracted OARC-provided DITL data and analyzed it based on the structure outlined in Appendix B to this letter. The slight differences between the NTAG analysis and Interisle’s are discussed in Appendix C.
- NTAG members convened a live TLD Security Forum on 22 August to discuss the merits of the Interisle report and how we can supplement it with what we believe is more reasonable analysis. Engagement has been so compelling that it has encouraged NTAG to move ahead on the efforts described below, as well as to convene another live TLD Security Forum in Washington, D.C. on 1 October. ICANN Staff and Board are more than welcome to attend and join the many other stakeholders who plan to participate.

We do not wish to toot our own horn, but need to stress how incredible it has been for NTAG to unite on the above efforts within such a brief period of time (i.e., 20 calendar days since staff posted name collision for public comment). You can imagine how daunting it has been to gather and analyze data that helps to provide perspective to the Interisle study and Verisign’s comments. Interisle has been commissioned to conduct the name collision study for ICANN. Verisign, as a root server operator, has private access to data. As such, both entities have had at least a six-week head start to access and

1 https://tldssecurityforum.eventbrite.com
analyze root query data and frame discussions about name collision and security risks with the new gTLD program.

Based on the NTAG’s efforts to date, below we outline as objectively as possible what we believe could sufficiently mitigate name collision risks and why. We believe that this letter can help the Board with its effort to reach a balanced (i) understanding of cyber security impacts with the new gTLD program and (ii) decision of the name collision issues.

Massive Overcounting of Collisions

Our initial response letter pointed out an obvious issue with using root query data as a proxy for measuring the proportion of DNS queries made against non-delegated TLDs: caching. The caching issue overestimates the number of collisions as a fraction of overall query traffic in multiple ways, due to both the effect of delegated domains having long TTLs and the standard treatment of TTLs on negative responses by recursive DNS servers.

It is our understanding that Google intends to submit a detailed paper on this issue, based upon data collected from their recursive PublicDNS system. We urge the Board and Staff to pay close attention to their findings, which include the fact that the root data set often overstates the number of collisions by a factor of 39x (.corp) and even up to 500x (.ice).

Our Suggested Mitigations

The NTAG has come up with a set of mitigations that we believe reduces the risk from name collisions to an acceptable level, and to well below the risk of any other TLD delegated in the last decade.

The mitigations fall into two categories: those applied to all TLDs and those applied to only the most problematic.

We recommend that ICANN mandate implementation of the following protections at the registry level for each new TLD:

- Prevent registration of second level domains corresponding to the major existing gTLDs2.
- Prevent registration of specific terms related to the Internet, networking, protocol, web related file extensions or common internal hostnames. This list is included in Appendix D.
- Exclude terms that a rightful owner has entered into the TMCH, require a one-year hold on second-level domain registration for strings that present more than a certain number of queries represented in the DITL data, with a custom hold list generated for each proposed TLD. In Appendix A, NTAG presents a chart showing collision risks before and after applying the proposed hold for strings that present more than 10,000 queries. The 10,000 threshold is merely a suggestion. We leave it to the Board and Staff to decide the acceptable risk threshold. In a

---

2 com, net, org, biz, edu, gov and mil.
couple of weeks, NTAG will be able to generate and report a custom list for each proposed new TLD at various threshold levels: 50,000 queries, 20,000 queries, 10,000 queries, 5,000 queries, and 1,000 queries. An example list for one proposed new TLD, .med, is included in Appendix E, along with the list of all 210 strings that would exceed a 10,000 query threshold.

- Block the registration of the decimal integers between 0 and 255 as an SLD under the new TLDs. This will prevent collisions from malformed reverse DNS lookups on IPv4 addresses.

The NTAG believes that this is a good base set of requirements, but that all of these protections could be reconsidered on a case-by-case basis using the RSEP process. It is likely that most of these protections would not be relevant to closed ".brand" TLDs, and we look forward to discussing more nuanced protections for those applications.

For three of the most controversial new TLDs, .corp, .home and .mail, we propose the following path forward:

1. That string contention be resolved in the normal order of business for all three strings.
2. That contracting and delegation move forward with the parties that emerge from contention, assuming that they agree to the following protections:
   a. That they operate a 90 day sunrise period for all strings.
   b. That they operate a one-year extended sunrise period for the most common brand names seen globally in name collisions. A tentative list of about 800 brand-related names, generated from the top 2000 strings seen as SLDs, is available in Appendix E.
   c. That they operate a one-year hold for the top 50 SLDs seen in requests for each TLD, assuming such strings have not already been included in previous lists. A list of these names is found in Appendix G.
   d. That while domains could be pre-sold for each TLD that no second level names will be delegated for a period of one year.
   e. That the TLD server logs from each of these new TLDs be turned over to OARC monthly for analysis by interested parties. This would allow ICANN and others to gauge the effectiveness of this risk mitigation program as it operates.
   f. That these protections could be shortened via the RSEP process, if supported by DNS query data.

Conclusion

In summary, our research shows that only a few known items are outstanding and require one or two weeks of further study. NTAG will follow-up this letter with a final report that includes these remaining items. Notably, NTAG will provide a report of registration hold lists for all proposed new TLDs and charts showing how implementing such holds would reduce name collision queries below certain risk thresholds, including the threshold which Board and Staff are to determine represents an adequate level of risk. What is known now provides strong support for permitting the new gTLD program to safely and successfully proceed, provided NTAG’s proposed mitigations are implemented.
Safety and success of the new gTLD program is important and achievable. Complaints with one new gTLD could risk prospects for all participants, inside and outside of ICANN. Because NTAG members understand this obvious correlation between consumer trust and new gTLD adoption, we’re committed to taking actions that align with the interests of Internet users. Name collision mitigations are but one example where we strongly believe that launch of the new gTLD program could help to enhance cyber security for users.

To help Board and Staff develop a fuller view and understand our commitment to cyber security, we encourage attendance at the second NTAG-sponsored TLD Security Forum. Specifically to Board members, irrespective of the stakeholders or committee(s) you represent, we believe that participation in the forum could help to inform your decision for moving forward on the name collision issue and deepen your understanding of cyber security for related future issues. Simply RSVP your interest in attending the TLD Security Forum at https://tldsecurityforum.eventbrite.com.

Among other content, a live overview of the issues — including an update of the remaining research items — will be provided. Participants can also look forward to hearing from voices outside of the ICANN community with a relevant take on Internet security issues. These conversations continue the productive talks from the first forum, where experts and stakeholders concerned with cyber security, including PayPal and a major certificate authority provider, joined the ICANN community to help address name collision risks, promote security with the new gTLD program and otherwise provide input on the evolution of the Internet as championed by ICANN.

Thank you for considering the data and analysis in this letter. We appreciate you taking time to hear our views before deciding this issue, especially because we’ve only recently gained access to the data necessary to add valuable input to the discussion.

Sincerely,
Members of the NTAG
Appendix A: Before and After Mitigations

The NTAG modeled the effect of removing SLDs that cannot be registered (such as _msdcs), filtering for the Chrome random 10 letter domains (which account for no security risk) and applying our suggested mitigations against the 2013 DITL dataset. This included the special treatment (a one-year hold with exceptions for TMCH terms) of all SLDs that have received more than 10,000 requests in that dataset. The 10,000 request line is completely arbitrary, and was chosen by the NTAG as a strawman proposal. We will provide the Board and Staff with simulations of other thresholds in our next report.

*Figure 1: This chart shows the number of requests for the top-10 proposed TLDs in the DITL 2013 data, alongside the remaining collisions after taking into account the protections NTAG recommends in this letter.*
Figure 2: This chart shows the same information for the 10 most frequent TLDs outside of .corp and .home.
Figure 3: This chart shows the same before and after counts for the 200 TLDs following those in Figure 2.
Appendix B: Data Extraction Process

As part of the DITL (Day In The Life) project, data captures are provided to the DNS-OARC project. Participating root-server operators (and other DNS server operators) provide collections of compressed network traffic capture files. These captures are in a common format for network captures called PCAP that can be read by many different tools.

The root servers see queries for all sorts of DNS lookups, including many valid and invalid queries. The input dataset for 2012 is 5.21 TB compressed (for root servers A, C, E, F, H, I, J, K, L and M) and 2013 is 1.76 TB compressed (for root servers A, C, D, E, F, H, J, K, L and M). The 2012 data was captured between April 17th and April 19th, 2012. The 2013 data was captured between May 28th and May 30th 2013. Processing, decompressing and filtering these very large datasets takes considerable computing power. In order to make it easier for researchers and analysts to work with the data, it was decompressed, processed, filtered and converted to an easy to process form.

We used a tool with a long pedigree called "tcpdump" to process the PCAP files prior to filtration and conversion to intermediate forms. Tcpdump has been available since 1987 and is actively maintained to this day. Additionally, it is widely available for most platforms, including Linux, Mac, FreeBSD, Windows.

While decompressing and using tcpdump to convert the PCAP files, we created a collection of intermediate files that contain only the TLD strings of interest. The TLDs of interest included all applied for gTLDs and as well as a few other strings.

Subsequent to creation of the intermediate files, we produced per-TLD files. To create these, we aggregated the filtered intermediate files, and split them into one file per TLD. Each of these per-TLD files are sorted by SLD for easier subsequent analysis.
Appendix C: Discrepancies versus Interisle’s report

While preparing data for analysis, a different set of tools were utilized than those used for the Interisle report. This was a conscious choice, as it allowed us to work towards reproducing Interisle’s numbers without following the exact same methodology.

Interisle’s report utilized a customized version of PacketQ to query the DITL data, whereas we used tcpdump and a series of data processing scripts. Because of the different tools and methodologies, we did encounter differences in our results, but are able to explain most of the differences.

We processed the data for 2013 and 2012 separately, and had managed to get our total query counts for 2013 within 1.25% after correcting for a few early processing errors. Such a small difference could have been explained by rounding errors and small differences in how PacketQ and tcpdump process data.

It wasn’t until we examined the 2012 numbers and saw that our 2012 data was under-reporting by 21.21% that we discovered a bug in our data extraction and filtering phase: Invalid queries containing spaces were not included in our processed datasets! Fortunately, this only had a negligible effect on the processed 2013 data, as there were relatively few queries with spaces in 2013. Including queries with spaces brought the 2013 query count total up 1.36% more queries than Interisle’s report. The only TLD that was significantly impacted by this bug was .APP, which was under-counted by 327,000 invalid queries (20.44%).

The situation for 2012 was quite different as 2012 has a significant number of queries for “File moved-http://…” which contains spaces, as Interisle’s report shows, which they report made up about 18% of the 2012 data. We found there to be at least 168,778,000 or 22.3% queries with spaces in them in 2012. This brings our total for 2012 to 2.04% shy of Interisle’s counts for 2012.

While investigating differences in how PacketQ and tcpdump process DNS packets, a couple of observations were found that may help explain differences. Due to limited time, we were not able to thoroughly investigate these. The key take-away is that the DITL captures contain a large volume of invalid or malformed DNS packets, which may be processed differently by different software. The tools, however, produce very similar numbers, if not the same, when processing properly formed, valid queries.

---

3 Interisle’s report rounds figures to the nearest 1000, which can produce significant percentage differences when comparing small query volumes.
4 Interisle Consulting Group: Name Collision in the DNS, Table 10, Page 47.
5 Interisle Consulting Group: Name Collision in the DNS, Table 11, Page 47.
● Unless PacketQ is told not to limit queries to “questions”, if TCP queries are present (such as EDNS0), these may be included in query counts. This would have a very small effect of having PacketQ count more packets than tcpdump, which does not include the query string for TCP responses.

● When the PUSH flag is present on a query (TCP), PacketQ appeared not to count it, but tcpdump did report it. This too represented a small number of queries.

We are satisfied that both Interisle's and our own processing of the DITL data are valid, as our datasets result in numbers that are acceptably similar. Given the completely different analysis toolset and large quantity of data and short timeline, it is not feasible to work to explain every little difference in the numbers. At the macro level, the results are sufficient identical, and validate each other.
Appendix D: Reserved Technical Labels

root
wpad
isatap
host
intranet
jpg
wsf
dns
lan
undefined
autodiscover
pop
smtp
gif
printserver
css
comm
imap
file
anonymous
sip
sipexternal
sipinternal
internet
xml
null
pop3
hotspot
html
router
internal
msdcs
udp
tcp
Appendix E: Reserving SLDs with over 10,000 Collisions

The following is the list of TLDs found to have at least one valid SLD with more than 10,000 collision occurrences in the 2013 DITL data set. We propose that each new TLD registry reserve all SLDs with more than 10,000 collisions for an extended period.

<table>
<thead>
<tr>
<th>active</th>
<th>dish</th>
<th>legal</th>
<th>school</th>
</tr>
</thead>
<tbody>
<tr>
<td>ads</td>
<td>doha</td>
<td>link</td>
<td>schule</td>
</tr>
<tr>
<td>africa</td>
<td>earth</td>
<td>live</td>
<td>search</td>
</tr>
<tr>
<td>amazon</td>
<td>eco</td>
<td>llc</td>
<td>security</td>
</tr>
<tr>
<td>anz</td>
<td>ecom</td>
<td>lpl</td>
<td>services</td>
</tr>
<tr>
<td>aol</td>
<td>energy</td>
<td>lol</td>
<td>ses</td>
</tr>
<tr>
<td>app</td>
<td>ericsson</td>
<td>london</td>
<td>sex</td>
</tr>
<tr>
<td>apple</td>
<td>exchange</td>
<td>lpl</td>
<td>sfr</td>
</tr>
<tr>
<td>art</td>
<td>faith</td>
<td>ltd</td>
<td>show</td>
</tr>
<tr>
<td>auto</td>
<td>family</td>
<td>madrid</td>
<td>sina</td>
</tr>
<tr>
<td>baidu</td>
<td>farm</td>
<td>mail</td>
<td>site</td>
</tr>
<tr>
<td>bank</td>
<td>fashion</td>
<td>maison</td>
<td>sky</td>
</tr>
<tr>
<td>bar</td>
<td>fish</td>
<td>man</td>
<td>smart</td>
</tr>
<tr>
<td>bbc</td>
<td>flickr</td>
<td>management</td>
<td>sohu</td>
</tr>
<tr>
<td>bcn</td>
<td>foo</td>
<td>map</td>
<td>srt</td>
</tr>
<tr>
<td>beer</td>
<td>forum</td>
<td>matrix</td>
<td>star</td>
</tr>
<tr>
<td>bet</td>
<td>fox</td>
<td>med</td>
<td>stc</td>
</tr>
<tr>
<td>bing</td>
<td>free</td>
<td>media</td>
<td>studio</td>
</tr>
<tr>
<td>bio</td>
<td>gal</td>
<td>medical</td>
<td>style</td>
</tr>
<tr>
<td>blanco</td>
<td>game</td>
<td>meo</td>
<td>sydney</td>
</tr>
<tr>
<td>blog</td>
<td>games</td>
<td>mit</td>
<td>systems</td>
</tr>
<tr>
<td>bloomberg</td>
<td>global</td>
<td>mlb</td>
<td>taobao</td>
</tr>
<tr>
<td>bom</td>
<td>gmail</td>
<td>mnet</td>
<td>tata</td>
</tr>
<tr>
<td>boo</td>
<td>gmbh</td>
<td>mobile</td>
<td>tax</td>
</tr>
<tr>
<td>box</td>
<td>goo</td>
<td>moe</td>
<td>team</td>
</tr>
<tr>
<td>business</td>
<td>goog</td>
<td>mad</td>
<td>tech</td>
</tr>
<tr>
<td>bzh</td>
<td>google</td>
<td>music</td>
<td>telefonica</td>
</tr>
<tr>
<td>cal</td>
<td>green</td>
<td>naval</td>
<td>terra</td>
</tr>
<tr>
<td>cam</td>
<td>group</td>
<td>network</td>
<td>thai</td>
</tr>
<tr>
<td>caravan</td>
<td>health</td>
<td>new</td>
<td>top</td>
</tr>
<tr>
<td>care</td>
<td>here</td>
<td>news</td>
<td>toshiba</td>
</tr>
<tr>
<td>casa</td>
<td>home</td>
<td>nico</td>
<td>town</td>
</tr>
<tr>
<td>catholic</td>
<td>hospital</td>
<td>nyc</td>
<td>tube</td>
</tr>
<tr>
<td>cba</td>
<td>host</td>
<td>off</td>
<td>unicorn</td>
</tr>
<tr>
<td>center</td>
<td>hosting</td>
<td>office</td>
<td>vet</td>
</tr>
<tr>
<td>cisco</td>
<td>hot</td>
<td>one</td>
<td>video</td>
</tr>
<tr>
<td>city</td>
<td>hotmail</td>
<td>online</td>
<td>voyage</td>
</tr>
<tr>
<td>cloud</td>
<td>house</td>
<td>ooo</td>
<td>web</td>
</tr>
<tr>
<td>college</td>
<td>hsbc</td>
<td>orange</td>
<td>website</td>
</tr>
<tr>
<td>comcast</td>
<td>ibm</td>
<td>partners</td>
<td>weir</td>
</tr>
<tr>
<td>company</td>
<td>ice</td>
<td>philips</td>
<td>win</td>
</tr>
<tr>
<td>computer</td>
<td>ifm</td>
<td>photo</td>
<td>windows</td>
</tr>
<tr>
<td>corp</td>
<td>iinet</td>
<td>plus</td>
<td>work</td>
</tr>
<tr>
<td>cpa</td>
<td>ikano</td>
<td>porn</td>
<td>world</td>
</tr>
<tr>
<td>csc</td>
<td>imdb</td>
<td>prod</td>
<td>wow</td>
</tr>
<tr>
<td>data</td>
<td>inc</td>
<td>pub</td>
<td>xyz</td>
</tr>
<tr>
<td>dealer</td>
<td>ink</td>
<td>red</td>
<td>yahoo</td>
</tr>
<tr>
<td>delivery</td>
<td>institute</td>
<td>sap</td>
<td>yandex</td>
</tr>
<tr>
<td>dell</td>
<td>islam</td>
<td>sas</td>
<td>you</td>
</tr>
<tr>
<td>delta</td>
<td>itau</td>
<td>sbi</td>
<td>youtube</td>
</tr>
<tr>
<td>dev</td>
<td>kids</td>
<td>sbs</td>
<td>zone</td>
</tr>
<tr>
<td>digital</td>
<td>lanxess</td>
<td>sca</td>
<td></td>
</tr>
<tr>
<td>directory</td>
<td>law</td>
<td>scb</td>
<td></td>
</tr>
</tbody>
</table>
Appendix F: Suggested Reserved Brand Strings for .corp, .home and .mail

sap
compassgroup
adi
sanm
hospira
main
sealedair
worldnow
franke
abacus
delta
essilor
dci
davita
iusacell
astrium
advantech
sungardas
hologic
googlle
ticketmaster
directenergy
cencosud
mediatek
gmcr
yahoo
playtech
uff
blackbaud
checkfree
lpl
mgr
mapei
tomra
oldcastle
dunkinbrands
nbty
skype
iac
tkse
raag
lgdisplay
hager
albras
vodacom
bankcolombia
nmhg
bskyb
galpenergia
kuoni
networksolutions
fnac
leviton
cisco
elocoeingles
lan
cnooc
webtrends
mauser
senag
nexant
dcl
facebook
laerdal
masco
ray
luxottica
netservicos
smarttech
petroleum
intersil
music
iasis
wt
ex
telecomegypt
aeroflex
phh
asta
seal
bulgari
mcfar
arcelik
ahm
rminfra
peror
telekomaustria
youtube
netiq
you
mobinil
firmex
crucell
investec
google
grontmij
psdi
pmisi
cabelas
mexicana
trader
dell
aol
imtech
amx
qpass
nomos
kaynakholding
combiamovil
iplgrou
markets
noma
freshdelmonte
xtralis
kordsa
howdens
muhaldib
jenoptik
onenec
startups
chang
toshiba
att
allseas
kcedeitag
unitededex
sinopac
bcferries
gowlings
walbroom
econcern
netac
srh
gateway
impaq
atptis
banqueatlantique
mobity
thefa
stcolumba
pmi
dav
pvncdab
ccf
markit
sea
seal
alrajhi
zira
blooms
homeaway
pulse
blessedtrinity
asus
ityp1c
arcsot
dot
acer
medhost
jss
conres
china
solmelia
wirelessadvocates
wsp
aria
aries
moneos
rchn
navtech
skrivanek
adexus
libyana
tipper
mckoolsmith
psys
naes
publicisgroup
zentiva
alto
localiza
fack
groupalia
fcompany
sericol
sodimac
takemoto
mcap
rtlgrou
telesat
dhiglobalmail
eenterasys
hillstar
shopzilla
olysteel
sunday
kinepolis
ziemann
maxtaille
mechel
ht
colombia
twitte
youtub
fiorucci
mli
ccn
# Appendix G: Top 50 SLDs for .home, .corp and .mail

## .home

- intra
- https
- pro
- tracker
- br0
- user
- nohost
- iphone
- e
desktop
- windows
- dhcppc0
- admin
- invalid
mail
- images
- computer
gmail
- pc2
server1
- lifbcibllhkhosfpjfnlhfpfgn
- pldf1
- i
- laptop
- error
- videos
activationbbr
- n
- b
- servidor
- xxx
- icongo
- s
- namics
- ieeb
dhcppc1
- a
- ufi
- hotmail
- mybookworld
- wdtvlive
- ipad
- com0
- printer
- d
- user-hp
diskstation
- u
- php
- bit

## .mail

- system
- alico
- com
- net
- yahoo
- org
- g
- mail
- company
- hot
- army
- gateway
- navy
- google
- www
- receive
- local
- smtp
- gmail
- e
- aol
- hotmail
- embarq
- y
- pop
- rocket
- reidiff
- corp
- for
- cra
- professional
- usmc
- autodiscover
- android
- sazepad
- usarmy
- delete-me
- imap
- comcast
- upc
- jensen
- abwin
- att
- live
- gruppo+credemin
- metro
- win
- border
- company
- oma

## .corp

- lasercard
- csodandbox
- rel
- broker
- beauty
- rosinv
- amlaw
- vfa
- srggi
- grupoq
- domain
- uafc
dorseylaw
- orkianet
- hymall
tatenergo
eisenmann
- cvstarrco
dipcmi
csodmgmt
airwatch
cfs2
dealerdotcom
neos
- 247wpp
- imax
tecsys
cegos
meus
raccorp
fpprod
buyabs
bps
stibo
3pillar
canternet
ncic
grupomag
price
alior
ddefue001
associa
freetranslation
hmcmdomain
cardone
amscan
wrgd
all