An Analysis of new GTLD Universal Acceptance in the Web Environment

ICANN has commissioned APNIC Labs to investigate aspects of "Universal Acceptance" of new gtLD domain names, with particular focus on the visibility of names in the new gTLD zones in the Web environment.

APNIC Labs has been delegated 304 distinct sub-zones of generic top-level domains for the purposes of measuring overall and per-country acceptance of these domains by end users and the DNS system as a whole. While no systematic complete block of any particular top level domain across the entire Internet was detected using this measurement process, there are some indications that some Internationalized Domain Name (IDN) domains, and some specific domains in some countries appear to generate operational anomalies.

ICANN has requested APNIC Labs to explore an aspect of "Universal Acceptance" by testing the *visibility* of names within certain new Generic Top Level Domains (gTLDs) by end users. In this context *visibility* of a domain name is both a function of the DNS (Can names within a GTLD be successfully resolved by users' DNS resolvers?) and a function of the web (Can URLs using names within a GTLD be fetched by users' browsers?). A combination of DNS resolution operations and HTTP fetches are performed by the APNIC Labs measurement system, and this permits both visibility considerations to be tested for a collection of names from within a set of new gTLDs by a large selection of users worldwide.

The Test Process

Between the 9th of June 2015 and the 10th of July 2015, APNIC ran an online advertising campaign using Google Ads. The advertisement material contained an embedded script that included the directive to fetch a small set of set of URLs. Each URL referenced a simple 1x1 pixel "blot".

The DNS names used in these URLS were drawn from sub domains of 304 distinct sub-zones of generic top level domains. The URL contained a component in the DNS name that was a uniquely generated string, to ensure that the resolution of the DNS name and the subsequent fetch of the referenced object was not locally trapped by caches or other forms of intercepting middleware. Both the authoritative server for the DNS names, and the HTTP server for the URLs were operated by APNIC, allowing us to use the servers' logs to examine the user's actions in retrieving the test URLs.

The test run encompassed a total of 184,031,694 tests, involving 36,641,711 end users.

The test script embedded in the ad requests the retrieval of five URLs, each involving a unique domain name within each of 5 distinct gTLDs drawn from the larger set of gTLDs being tested. The test also involves a "result" fetch, which is executed either when the other five URLs have been retrieved or when a 10 second timer has elapsed. The "result" fetch includes additional arguments in the fetch string which are the encoded client-side measurements of the elapsed time taken to retrieve each of the URLs, or a "null" value if the URL was not retrieved when the 10 second timer occurred. In this test Adobe Flash was used as the scripting language.

As we are looking for patterns of "missing" fetches we have used the "result" fetch as the filter for accepting data from an individual experiment. We have excluded from consideration those instances where the user's system did not complete execution of the test script, in that they failed to return an explicit end-user result marker. This excluded some 77,814,048 domain tests from 15,398,181 users.

Total Users Tested:	36,641,711
Total Domain Tests:	184,031,694
Result-confirming Users:	21,243,530
Result-confirming Tests Applied:	106,217,646

(This level of attrition of 42% in not generating a result fetch is not unusual when the requirement of the measurement methodology requires the user to let the ad run until completion or 10 seconds.)

If there is systematic blocking of the visibility of a name in the context of its use in a URL, it may occur in a number of places.
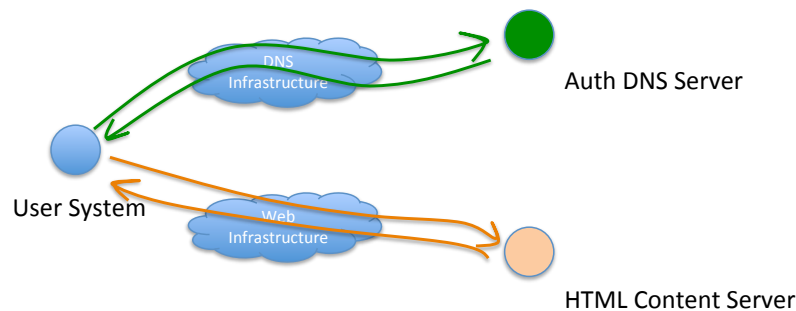
- DNS queries for a name may be blocked, in which case the authoritative name servers for the zone will not be passed any DNS queries from the network where the block is being performed.

- A more subtle form of blocking is where the DNS response is used to perform the block, either by substituting a synthetic "no such name" response or by rapidly responding with a synthetic response, so that the resolver will reject the subsequent genuine response as an unsolicited echo. In this case the authoritative name server will still see the DNS queries, but there will be no subsequent fetch of the Web object.

- Another form of blocking is to leave the DNS query and response alone but perform blocking on the Web object fetch. This can be performed by blocking the HTTP web query form reaching the web server. The server infrastructure also sees this as a a DNS query but no matching subsequent Web query.

- The last potential case is blocking of the delivery of the Web response to the client. This is not directly detectable by the server infrastructure, but can be inferred as a possible reason in those cases where the web server reports a URL query and response, yet the user's result report reports that it did not retrieve the URL.

Methodology

The set of gTLDs, and the names within these GTLDs to test was nominated by ICANN. For each one, a distinct sub-delegation was registered by ICANN, pointing to a second or third-level delegation (depending on the gTLD or ccTLD policy).

The set of registered sub-domains was as follows:

icann.academy	icann.accountants	icann.actor	icann.agency	icann.airforce	icann.army
icann.associates	icann.attorney	icann.auction	icann.audio	icann.band	icann.bar
icann.bargains	icann.beer	icann.best	icann.bid	icann.bike	icann.bio
icann.black	icann.blackfriday	icann.blue	icann.boutique	icann.brussels	icann.build
icann.builders	icann.business	icann.buzz	icann.bzh	icann.cab	icann.camera
icann.camp	icann.capetown	icann.capital	icann.cards	icann.care	icann.career
icann.careers	icann.casa	icann.cash	icann.catering	icann.center	icann.ceo
icann.cheap	icann.christmas	testo.church	icann.city	icann.claims	icann.cleaning
icann.click	icann.clinic	icann.clothing	ekzameno.co.il	icann.coach	icann.codes
icann.coffee	icann.cologne	icann.community	testo.company	icann.computer	icann.condos
icann.construction	icann.contractors	icann.cooking	icann.cool	icann.country	icann.credit
icann.creditcard	icann.cricket	icann.cruises	icann.cymru	icann.dance	icann.dating
icann.deals	icann.degree	icann.delivery	testo.democrat	icann.dental	icann.dentist
icann.desi	icann.diamonds	icann.diet	icann.digital	icann.direct	testo.directory
icann.discount	testo.domains	icann.durban	icann.education	testo.email	icann.energy
icann.engineer	icann.engineering	icann.enterprises	icann.equipment	icann.estate	icann.events
icann.exchange	icann.expert	testo.exposed	testo.fail	icann.farm	icann.fashion
icann.finance	icann.financial	icann.fish	icann.fishing	icann.fitness	icann.flights
icann.florist	icann.flowers	icann.forsale	icann.foundation	icann.frl	icann.fund
icann.furniture	icann.futbol	icann.gallery	icann.garden	icann.gent	icann.gift
icann.gifts	icann.gives	icann.glass	icann.graphics	icann.gratis	testo.gripe

icann.guide	icann.guitars	icann.guru	icann.haus	icann.healthcare	icann.help
icann.hiphop	icann.holdings	icann.holiday	icann.horse	icann.host	icann.hosting
icann.house	icann.how	icann.immo	icann.immobilien	icann.industries	icann.ink
icann.institute	icann.insure	icann.international	icann.investments	icann.jetzt	icann.joburg
icann.juegos	icann.kaufen	icann.kim	icann.kitchen	icann.kiwi	icann.koeln
icann.land	icann.lawyer	icann.lease	testo.lgbt	icann.life	icann.lighting
icann.limited	icann.limo	icann.link	icann.loans	icann.london	icann.maison
icann.management	icann.market	icann.marketing	icann.media	icann.memorial	icann.menu
ekzameno.mobi	icann.moda	icann.moe	icann.money	icann.mortgage	trial.moscow
icann.nagoya	ekzameno.name	icann.navy	icann.network	icann.ngo	icann.ninja
icann.okinawa	icann.ong	icann.ooo	icann.osaka	icann.paris	icann.partners
icann.parts	icann.party	icann.photo	icann.photography	icann.photos	icann.pics
icann.pictures	icann.pink	icann.pizza	icann.place	icann.plumbing	testo.poker
icann.press	icann.productions	icann.properties	icann.property	icann.pub	icann.qpon
icann.quebec	icann.recipes	icann.red	icann.rehab	icann.reisen	icann.rentals
icann.repair	icann.report	testo.republican	icann.rest	icann.restaurant	icann.reviews
icann.rip	icann.rocks	icann.rodeo	icann.ryukyu	icann.saarland	icann.sarl
icann.schule	testo.science	icann.services	testo.sexy	testo.shiksha	icann.shoes
testo.singles	icann.social	icann.software	icann.solar	tryout.solutions	icann.soy
icann.space	icann.supplies	icann.supply	icann.support	icann.surf	icann.surgery
testo.sx	tryout.sx	icann.systems	icann.tatar	icann.tattoo	icann.tax
testo.technology	testo.tel	icann.tienda	icann.tips	icann.tires	icann.tirol
testo.today	icann.tokyo	icann.tools	icann.town	icann.toys	icann.trade
icann.training	icann.university	icann.uno	icann.vacations	icann.ventures	icann.vet
icann.viajes	icann.villas	icann.vision	icann.vlaanderen	icann.vodka	icann.voting
icann.voyage	icann.wales	tryout.wang	icann.watch	icann.webcam	icann.website
icann.wedding	icann.whoswho	icann.wien	testo.wiki	icann.work	icann.works
icann.world	testo.wtf	testo.xyz	icann.yoga	icann.yokohama	icann.zone

icann.xn--6qq986b3xl	xn--80akhbyknj4f.xn--80adxhks
xn--80akhbyknj4f.xn--80asehdb	xn--80akhbyknj4f.xn--80aswg
xn--80akhbyknj4f.xn--c1avg	xn--11b5bs3a9aj6g.xn--i1b6b1a6a2e
xn--hgbk6aj7f53bba.xn--mgbab2bd	xn--hgbk7fnr.xn--ngbc5azd
xn--0zwm56d.xn--nqv7f	icann.xn--q9jyb4c

The registration for "testo.tel" was not successful. However tests were made against this name. This stands as a 'control fetch' of a non-delegated domain and should show for any sub-sample of our experiment a complete failure to fetch the test in DNS and the Web.

The domain name "eczameno.co.il" was included as a check of what has been commonly assumed to be the existence of systematic filtering of the ".il" ccTLD by some countries or networks.

Two subdomains were included from the ccTLD ".sx" being "testo.sx" and "tryout.sx" –which caused approximately double the sample rate of other gTLD/ccTLD.

The authoritative name servers of these listed delegations was assigned to an IPv4 and IPv6 address of a DNS and Web server operated by APNIC Labs in a data centre located in Dallas, Texas. Each zone was declared to exist with this server as the sole authoritative server for the zone. The zone was populated with a wildcard entry, and unique name label elements were included in every test. This was to ensure that that the corresponding DNS address resolution queries for any domain name would be answered by the server, as there was no prior cached entry for each unique name. The address record of the name was again that of an APNIC server located in the same data centre, so any consequent activity, such as the fetch of a web object, would also be directed to an APNIC server.

APNIC Labs configured its measurement system to proffer 5 selections from the list of names to each client performing the measurement within the impression of each advertisement. The 5 names were taken in sequence from the list, modulo the list size so that each client was asked to fetch 5 distinct named entities drawn from the entire list of delegated names.

Each client is selected by virtue of the measurement script being embedded in an online advertisement system (operated by Google), so that the client selection and subsequent measurement activity is based on advertisement impression. This resulted in an overall test volume for each of the 304 names by approximately 300,000 clients in this 30 day interval.

Data Analysis

We start by looking at all experiments that generated a result record that was seen at the server. This result record listed the outcome of the attempts to fetch 5 URLs within 10 seconds, where for each URL the result either noted the client's measured elapsed time to fetch the URL if successful, or a null result if the fetch was not recorded by the client.

We then look through the Web Server logs to see if there are instances of web fetches seen at the server that are not logged in the result record. By comparing the time of the web fetch with the time that the result was received then we can classify this as being either a "late" web fetch (where the 10 second script timer at at the client side expired before the set of individual tests was complete), or a more complex case where while the server believed that it had delivered the requested object within the time parameters of the experiment, the user's browser did not acknowledge its receipt.

We then look through the DNS logs to see if there are instances of DNS queries for objects which were not fetched by the client.

For each experiment there are four potential outcomes:

1. *Web* where the client lists the object as received in the results report.

2. *No DNS* where the DNS server has not seen any queries for the object.

3. *No Web* where the web server has not seen any fetch of the object.

4. *Blocked Web* where the web server has recorded the delivery of the web object to the client, but the client is not listing the object as received, and the recorded web fetch happened 2 seconds (or more) earlier than the result record.

5. *Late Web* where the web server has recorded the delivery of the web object to the client, but the client is not listing the object as received, and the recorded web fetch happened after the result record.

The results are aggregated by country of origin of the user's IP address, as reported by the Maxmind IP-to-country geolocation database.

Results

Each name was tested by a minimum of 299,000 users. The aggregate result for each name is shown in Figure 1.
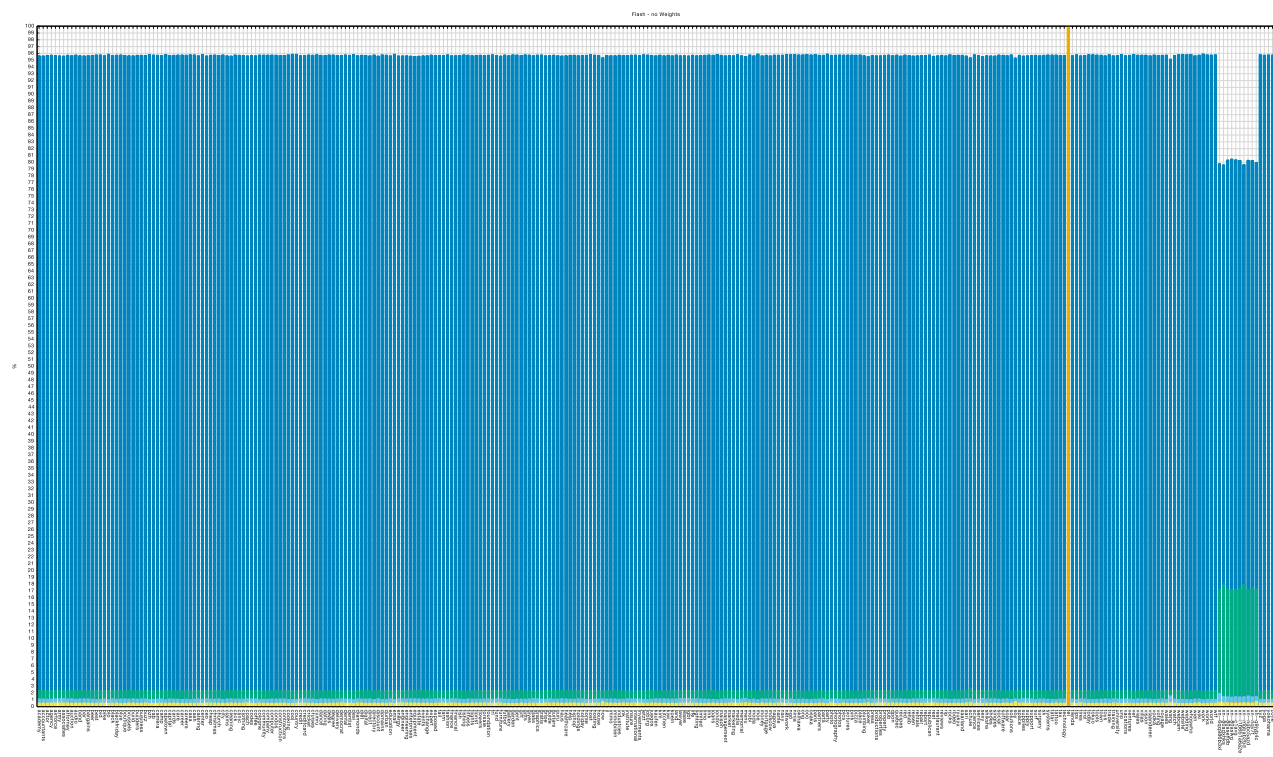

Figure 1 – Overall Experiment results for IDN names.

With the exception of the 10 punycode-encoded IDN names, and the undelegated .testo domain, all other domains recorded a very similar response at an internet-wide level.

These names had a result profile of:

1. *Web* success rate of 96%

2. *No DNS* rate of 0.5%

3. *No Web* rate of 1%

4. *Blocked Web* rate of 2%

5. *Late Web* rate of 0.5%

There is perhaps a question to be asked here as to why, across all these 294 names, was the result so consistent and, relative to our expectations, so bad? Is there really a failure rate of around 5% of URL queries where the name is unique? Is the web environment now becoming highly dependent on various intermediaries to operate reliably in all cases? This question is perhaps outside the scope of the commissioned study, and is worthy of further study to analyze the web failure profile exposed in this activity..

The 10 punycode-encoded names had a response profile of:

1. *Web* success rate of 80%

2. *No DNS* rate of 0.5%

3. *No Web* rate of 1.5%

4. *Blocked Web* rate of 17%

5. *Late Web* rate of 0.5%

There is clear evidence of some issue with the Web fetch for punycode-encoded IDN names. It is unlikely that there is a systematic block of such names in HTTP middleware. The problem appears to be within the client environment and a possible explanation is that while the script being executed within the browser resolves the DNS name and performs a web fetch of the object, the script does not appear to recognize the Web object as the requested object within certain browsers. This may be due to the internal framework of the Flash engine in some browsers performing a translation of the Punycode-encoded name into its Unicode equivalent, and then failing to recognize the returned named object as being the originally requested object due to the internal name translation.

We analyzed the distribution of browsers in this experiment by parsing the User Identity string presented to the Web Server in the HTTP protocol as part of the Web fetch. The measured distribution of browser use in this experiment was:

Chrome	78%
Firefox	12%
MSIE	6%
Opera	3%
Safari	1%

The profile of behavior of the fetch of punycode-encoded names is shown in Figure2
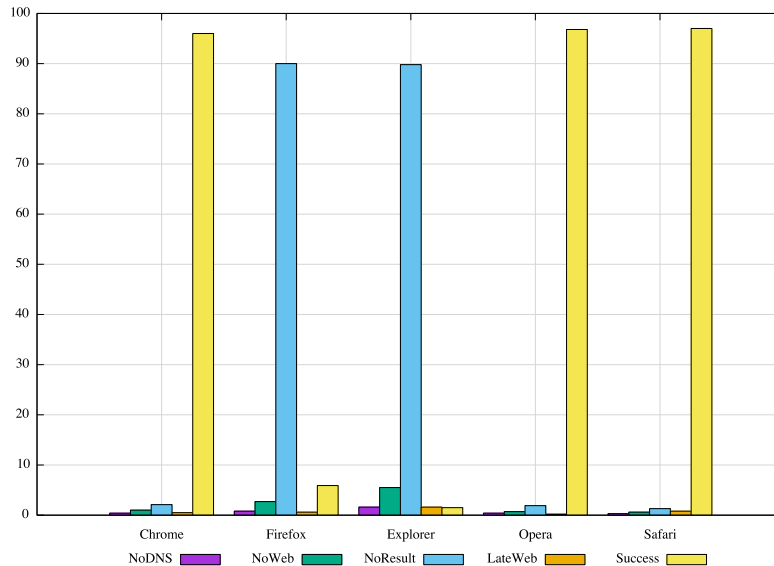
Figure 2 – Result Profile by Browser for Punycode-encoded IDN names.

It is evident that the difference in behavior for the punycode-encoded IDN names in this experiment is due to some behavior in the Firefox and Explorer browsers. This may be due to the browser itself, or possibly due to the Adobe Flash engine used by the browser, as this measurement uses a Adobe Flash as its script engine, or the interaction between the Flash engine and the browser platform.

To determine whether the problem lies in the browser or in the Flash implementation within the browser environment we set up a second test using HTML5 as the script engine, testing four names rather than the complete set. The results of this second experiment are shown below:

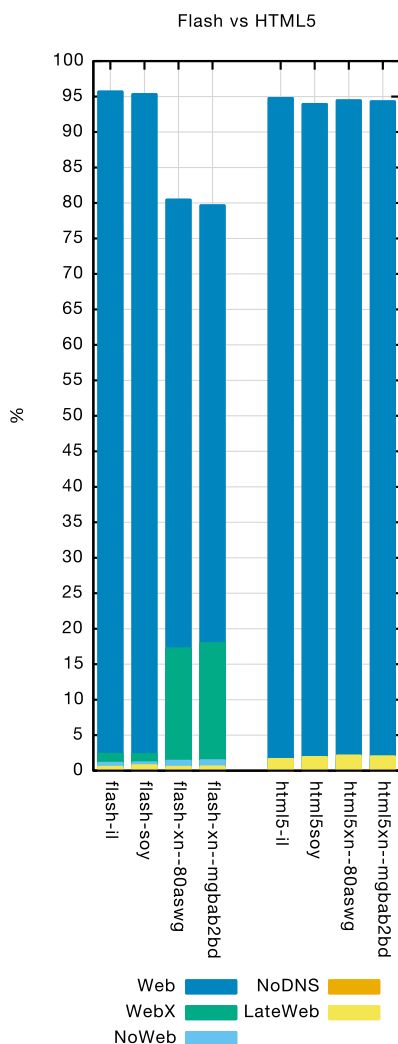| gTLD	Script	Experiments	No DNS (%)	No Web (%)	X Web (%)	Late Web (%)	Success
.il	Flash	903,019	0.4%	1.1%	2.3%	0.5%	95.7%
	html5	1,849,583	0.5%	1.5%	1.6%	1.6%	94.8%
.soy	Flash	299,628	0.5%	1.1%	2.3%	0.7%	95.3%
	html5	1,849,583	0.9%	1.7%	1.5%	1.90%	93.9%
.xn--80aswg	Flash	298,955	0.5%	1.4%	17.2%	0.5%	80.4%
	html5	1,849,583	0.5%	1.5%	1.5%	2.1%	94.4%
.xn--mgbab2bd	Flash	708,818	0.5%	1.4%	17.9%	0.5%	79.6%
	html5	1,849,583	0.7%	1.5%	1.5%	2.0%	94.3%

Figure 3 – Comparison of Result Profile by Script Engine for Punycode-encoded IDN names.

The absence of any anomalous results using HTML 5 and Punycode-encoded IDN domain names points in a rather compelling manner to problems with the combination of the Flash engine and Firefox and Explorer browsers, where is appears that the Flash script appears to lose track of the distinction between the Unicode name and its punycode-encoded equivalent.

The second phase of the analysis was to determine if there were any systematic problems with the web access to URLs that used DNS names within this set of gTLD names. An annex to this report contains both the data on a per country basis and a plot of the observed behaviours of each of the 304 names. What we looked for were situations where a name experienced a visibility problem for most the users within a particular economy.

Within the set of tested names we noted 3 instances where this appears to be happening.

A name within *.sexy* may be blocked in Iran. What we observed here is that the users were able to resolve the name using the DNS, and generate a HTTP fetch for the referenced URL. However the resultant 1x1 blot was not passed back to the end user, as no users reported that they received the URL in the scope of this test. However the sample count was 30 in this case, so this is a somewhat qualified observation

IR	tests	No DNS	No Web	Web X	Late Web	Success	No DNS %	No Web %	Web X %	Late Web %	Success %
sexy	30	1	2	25	0	2	3.33%	6.67%	83.33%	0.00%	6.67%

The name *.il* may be blocked in Iran. The observed behavior was somewhat different to the previous case of *.sexy*, in that .il appears to be resolved normally for most users by the DNS, but no Web fetch was observed, indicating that a likely cause of the lack of visibility was in the interception of the DNS response, or the use of an HTTP proxy that is performing a filtering function. With 116 samples this is also a somewhat qualified observation. Of course .il is not one of the gTLDs per se, but was included in this experiment as a control point, as there were reported visibility issues with this name from some countries.

IR	Tests	No DNS	No Web	Web X	Late Web	Success	No DNS %	No Web %	Web X %	Late Web %	Success %
il	116	14	91	0	2	9	12.07	78.45	0.00	1.72	7.76

The name *.il* may be blocked in Syria. The observed behavior was similar to that observed in Iran where *.il* appears to be resolved by most users, but no Web fetch was observed, for similar reasons to those in Iran. Again, this is a relatively small sample count, so the observation is necessarily qualified.

SY	Tests	No DNS	No Web	Web X	Late Web	Success	No DNS %	No Web %	Web X %	Late Web %	Success %
il	110	11	81	2	0	16	10.00	73.64	1.82	0.00	14.55

There were four other instances of a failure rate of > 50% seen in the data, from names tested by users in Togo (TG) and Mauritania (MR). However, it is likely that given the small sample counts and the location of these users in Africa and the closest servers for the experiment in Europe, the reason why there were a high relative proposition of *Web X* ("Blocked Web") counts in these cases is due to high latency of the network path between the servers and the users undertaking the experiment

CC	gTLD	Tests	No DNS	No Web	Web X	Late Web	Success	No DNS %	No Web %	Web X %	Late Web %	Success %
MR	durban	15	1	3	4	0	7	6.67	20.00	26.67	0.00	46.67
TG	watch	11	0	1	5	0	5	0.00	9.09	45.45	0.00	45.45
TG	webcam	11	0	1	4	1	5	0.00	9.09	36.36	9.09	45.45
TG	website	12	0	1	6	0	5	0.00	8.33	50.00	0.00	41.67

Not all countries show the same relative counts between the five response classes. Some countries show consistency across all tested names, while others show a different response. The environmental conditions that have a bearing on this experiment include factors such as the network path delay between the user and the experiment's servers, the presence or otherwise of various forms of DNS and Web middleware (such as Web Proxies and similar), the relative level of performance of these middleware units and the level of congestive load experienced in the DNS and in the end-to-end data paths used by the HTTP fetch.

There are national populations that appear to be behind heavily loaded infrastructure, as evidenced by high Web loss rates and at times accompanied by high DNS loss rates. Some national data patterns appear to be indicative of Web proxy systems that impose additional latency in web fetching, while others show an elevated DNS loss rate but not Web loss.

However, the speculation as to the root cause of these differences is necessarily limited by the measurement system used, embedded Flash scripts in online ads, and the degree of variability in end user environments, and this is straying beyond the original objective to detect any levels of systematic blocking of the use of names in any of the new gTLS name spaces.

Conclusions

There is little evidence in the measurements conducted by APNIC Labs for ICANN of any form of widespread systematic blocking of particular gTLD names at a national level for this subset of recently introduced gTLD names for "conventional" (non–IDN) names.

The poor performance of punycode encoded IDN names appears to be a combination of the use of Adobe Flash as the script engine in combination with the Explorer and Friefox browser platforms. Using an HTML5 script appeared to eliminate the failure issues with these names..

One country, Iran, shows some evidence of a piecemeal block of Web names within the ".sexy" gTLD.