Root Cause Analysis - New gTLD Collisions

Introduction

In 2013, the International Corporation for Assigned Names and Numbers (ICANN) began allowing new top-level domains (TLDs) to be introduced into the DNS root zone. Analysis showed that this new practice might adversely affect existing networks and systems, because of *name collisions*: the notion that a system uses a given DNS namespace in *private* and *relies* on it not resolving in the public DNS, but then, through delegation, that namespace becomes publicly resolvable. Because of the potential problems associated with name collisions, newly delegated TLDs were required to go through a period known as "controlled interruption," beginning in August 2014. This practice, described in more detail hereafter, was intended to make users and administrators that *might* be affected by a TLD's delegation aware of its delegation preemptively—before the problems became critical.

ICANN's Security and Stability Advisory Committee (SSAC) commissioned the Name Collisions Analysis Project (NCAP) to "facilitate the development of policy on Collision Strings to mitigate potential harm to the stability and security of the DNS posed by delegation of such strings" (https://community.icann.org/display/NCAP/). This document is part of the NCAP effort. In particular, this study seeks to analyze various aspects of name collisions and controlled interruption since controlled interruption was instituted and to identify the root cause of related incidents reported by affected parties to ICANN. The analysis primary takes into considering TLDs delegated between August 2014 and June 2021. Three data sources are used in this analysis:

- collision reports submitted via ICANN's name collisions Web submission form;
- passive DNS from the 100 days of controlled interruption during the initial delegation of each TLD; and
- root query data from the 48-hour once-yearly day-in-the-life (DITL) collection from 2014 to 2021.

We begin with some technical background information related to our analysis and then briefly describe our data sets. We then perform an analysis of the name collision reports submitted to ICANN. Next we describe our methodology for quantifying the private use of newly delegated TLDs, and we share the results of our analysis of controlled interruption and leaked DNS queries intended for privately maintained namespace. We describe a survey that we commissioned to obtain more qualitative data associated with our analysis. Finally, we summarize our findings and propose future work.

Background

This section provides technical background related to our study.

DNS Suffix Configuration

The network configuration for most operating systems includes the option for a DNS domain (e.g., `example.com`) to be specified for various purposes. The system's stub resolver library, which is used by applications to resolve DNS names to addresses, might apply this domain to unqualified DNS names that are to be resolved (e.g., `foo` becomes `foo.example.com`). Or the domain might be used to identify certain network resources associated with the organization, such as the organization's HTTP proxy server (see section XX) or potential routers for IPv6-over-IPv4 tunneling (see section ??).

This domain is configured in the "domain" and "search" entries of /etc/resolv.conf on UNIX and Linux systems. In macOS, the DNS configuration pane contains a "Search Domains" box to add this domain. On Windows, the "DNS suffix search list" is used.

Throughout this document, we use the term *DNS suffix* to refer to this domain, independent of the specific system on which it is configured.

Controlled Interruption

Some systems inadvertently query the public DNS for names under a non-existent TLD, for a variety of possible reasons. *Prior* to the delegation of the TLD in the root zone, these names would not resolve but would rather result in an NXDOMAIN (name error)--or negative response. In some cases, a negative response from the public DNS was *relied on* to properly access a given resource (e.g., search list processing). In other cases, a negative response from the public DNS would simply *prevent* a system from accessing a given internal resource except from within the proper network for doing so (e.g., private namespace used within a corporate network). In all cases, negative responses played a role in *expected* behavior.

Controlled interruption involves inserting *wildcard* records in the otherwise empty zone file associated with a previously undelegated TLD. The wildcard A record in the zone file maps to a non-routable address: 127.0.53.53. Thus, *any* query made to the public DNS for names under that TLD result in a positive response—as opposed to the negative response that would have resulted prior to controlled interruption.

Controlled interruption has been required of all TLDs delegated in the root zone since August 2014, for the first 100 days of its delegation. In cases where negative responses were required for expected behavior, it was expected that systems encountering controlled interruption would experience some sort of disruption to their "normal" behavior, a sort of signal that something had changed in the public DNS. Additionally, it was the hope that this disruption would be noticed by the affected parties, such that they would investigate and take action, by reporting the problem and/or changing their configuration.

Chrome Browser NXDOMAIN Probing

On startup, the Google Chrome Web browser historically issued three queries, appending the system DNS suffix (see section ??) to three randomly-generated alphabetic strings. This is to detect infrastructure providing synthetic positive responses to DNS queries that would otherwise be classified as name errors (NXDOMAIN). During controlled interruption for a given TLD, queries under that TLD related to Chrome NXDOMAIN probing result in positive DNS responses.

WPAD-Related Queries

With the Web Proxy Auto Discovery Protocol (WPAD), browsers (e.g., Mozilla Firefox and Google Chrome) and operating systems (e.g., MacOS and Windows) auto-detect HTTP proxy settings using the DNS and HTTP. The specification designates that a WPAD client append the DNS suffix with which a system is configured (see section XX) to the label `wpad`. If no answer is found for the newly-formed domain name, then the left-most label in the DNS suffix is stripped, and `wpad` is prepended to the resulting suffix. Thus, a browser on a system configured with DNS suffix `foo.example.com` would issue a DNS query for `wpad.foo.example.com` then (assuming the domain name did not resolve) `wpad.example.com`, etc. This process is repeated until an answer is found or the suffixes are exhausted. During the controlled interruption period for a given TLD, all WPAD-related queries under the TLD result in positive DNS responses.

ISATAP-Related Queries

The Intra-Site Automatic Tunnel Addressing Protocol (ISATAP) is used for creating a link-local IPv6 address from an IPv4 address and discovering a neighbor through which IPv6 traffic might be tunneled. As part of this process, a host discovers potential routers by performing a DNS lookup for the qname formed by appending the system's DNS suffix (see section ??) to the string `isatap`. Thus, for a system configured with the DNS suffix `example.com`, the DNS lookup would consist of a lookup for `isatap.example.com`.

Data Sets

In this section, we describe the data sets that were used as the basis for our analysis.

Name Collisions Reports Submitted via ICANN's Web Form

After ICANN began introducing new TLDs into the root zone, a Web form was created whereby users could submit reports of problems experienced, each potentially related to the delegation of new TLDs (https://www.icann.org/en/forms/report-name-collision). Each report contained the date of the report, the TLD in question, a brief description of the problem, and contact information of the submitter. We use the data from these reports to better understand user and organization experience associated with the delegation of new TLDs in Section ??.

DNSDB

DNSDB, operated by Farsight Security, is a DNS database populated by passive DNS sensors at operators world-wide. It contains historical domain-name-to-resource mappings going back more than 10 years. For example, it could show that `example.com` (`A` record type) resolved to 192.0.2.1 from March 2014 to October 2015 and to 192.0.2.2 from December 2015 to February 2019. It also supports historical response data for other record types, including NS, MX, and others. However, it only contains an entry where there is a legitimate mapping observed by a sensor. Thus, there is no entry in DNSDB in either of the following cases:

- A mapping exists, but there was no sensor deployed to observe it; or
- The query for a given name results in a negative response (i.e., no mapping)

We used DNSDB to create two data sets in this work: *query names* observed *during* the controlled interruption period; and *mappings* observed *since* controlled interruption.

Controlled Interruption Queries

We used ICANN's published list of delegated strings (https://newgtlds.icann.org/en/program-status/delegated-strings) to obtain the list of TLDs delegated between August 2014 and June 2021, as well as the delegation date of each. August 2014 was when the requirement for controlled interruption began for newly delegated gTLDs. The following table shows the breakdown by year of each of the 885 domains delegated during the time period:

Year	**TLDs Delegated**	**Year**	**TLDs Delegated**
2014 (Aug - Dec)	131	**2018**	5
2015	390	**2019**	3
2016	340	**2020**	4

2017	12	**2021 (Jan - Jun)**	0
Total: 885			

For each of the new TLDs delegated, we issued a DNSDB query to solicit mappings observed during the dates of its control interruption period—i.e., the first 100 days of its delegation. Because controlled interruption results in a mapping (i.e., to 127.0.53.53) for *any* DNS queries under the TLD, the DNSDB queries effectively yielded every DNS name *queried* during the controlled interruption period, for DNS names under the new TLD, along with a count of how many times it was queried. We refer to this data set as DNSDB-CI.

Queries Post Controlled Interruption

Requesting a complete history of *all* DNS mappings observed for every one of the 885 new TLDs delegated since their controlled interruption period ended would have been infeasible because the data sets would be so huge. However, for this analysis, we were interested in only a subset of the namespace under each TLD—that associated with DNS suffixes, which are identified in section ??. Therefore, we sought the mapping history of 2,266 subdomains of the new TLDs. These subdomains represented 166 of the new TLDs and were selected using the methodology described in Section ??. We issued queries to DNSDB for all query names under each of the 2,266 subdomains (using a wildcard DNSDB query, such as `*.example.com` for the DNS suffix `example.com`), in each case requesting all mappings observed since the 100-day period of controlled interruption for the TLD associated with the subdomain.   We refer to this data set as DNSDB-PostCI.

DITL

Various DNS root server operators contribute to a yearly collection of 48 hours of DNS queries observed at the root server system. This collection is known as the “Day in the Life” or DITL collection and is sponsored by the DNS Operations, Analysis, and Research Center (DNS-OARC). For this analysis, we extracted query information for queries associated with the 2,266 subdomains that we identify in section ?? for every year between 2014 and 2021, inclusive, from root letters A, C, H, and J. This subset of four root letters was selected because each of these letters was available in each of the DITL years we were interested in. For efficiency, we utilized a two-step process:

1. Identify queries for query names ending with each of the 166 TLDs; and
2. Further filter queries to those with query names ending in any of the 2,266 suffixes.

The result was a list of tuples consisting of query name and IP address for every captured query that corresponded to any of the 2,266 subdomains for eight yearly 48-hour traffic collections between 2014 and 2021. We refer to this data set as DITL-2014-2021. This is further described in Section ??.

Name Collisions Report Analysis

We now analyze the reports submitted to ICANN via the Web form (see Section ??). We note that this data set has inherent bias in two ways. First, the submission of the report itself implies that a user or organization was impacted in some way. Second, the submission implies that they found the online form, which means that the question of the *effectiveness* of the controlled interruption IP address (127.0.53.53) to trace the problem to ICANN and the delegation of new TLDs cannot be evaluated; there is simply nothing in this data set to compare *against*. Later in the paper (see Section ??) we survey a general audience of network administrators as well as a targeted audience of organizations potentially affected by the delegation of new TLDs—a study without those same biases.

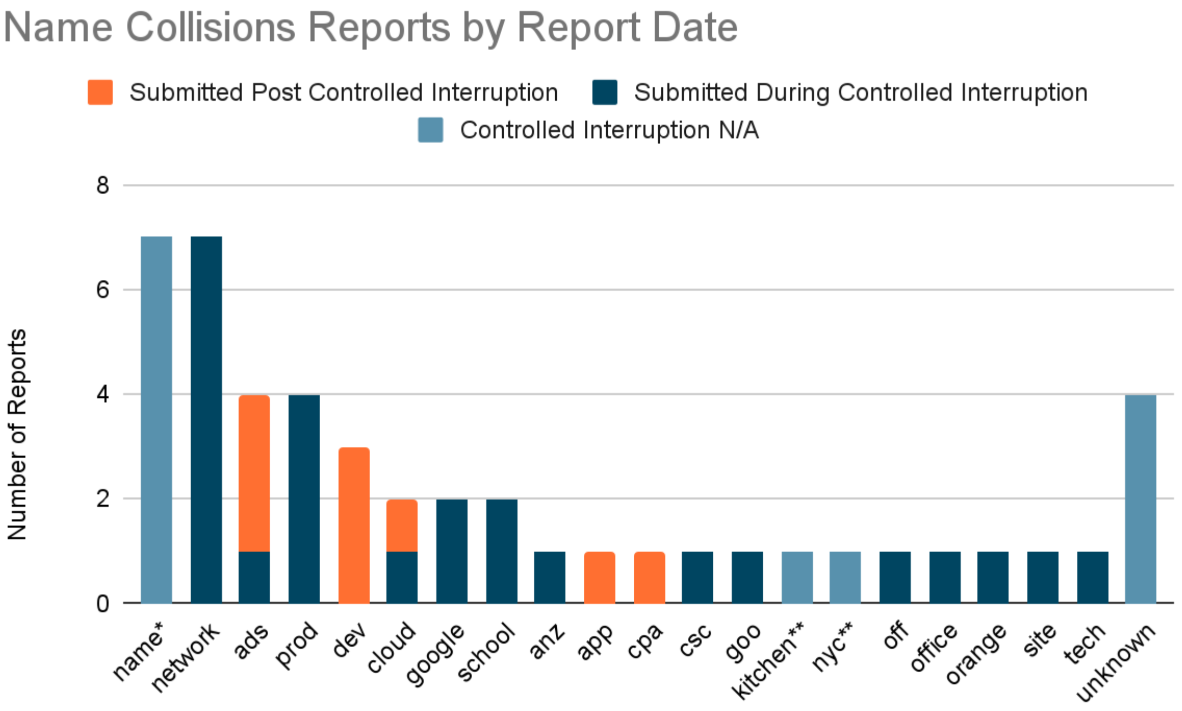
TLD Statistics

The following table contains a summary of the reports submitted, based on factors such as the date of the report, the TLD and its delegation date, and the reporting entity.

Total reports	**47**
do not include TLD	4 (9%)
include TLD	**43 (91%)**
delegated prior to new TLD program*	7 (16%)
delegated as part of new TLD program	**36 (84%)**
prior to controlled interruption (pre-Aug 2014)**	2 (6%)
with controlled interruption (Aug 2014 or later)	**34 (94%)**
report date is during controlled interruption	25 (74%)
report date is post controlled interruption	9 (26%)
reported by organization	24 (71%)
reported by individual	9 (26%)
reported origin unknown	1 (3%)
Total TLDs reported	**20**
delegated prior to new TLD program*	1 (5%)
delegated as part of new TLD program	**19 (95%)**
prior to controlled interruption (pre-Aug 2014)**	2 (11%)
with controlled interruption (Aug 2014 or later)	17 (89%)

While the table captures the data of all reports, we pay particular focus to the subset of 34 (72%) reports that pertain to TLDs delegated after the controlled interruption period. Of the 20 TLDs mentioned, 17 (84%) fit this category.

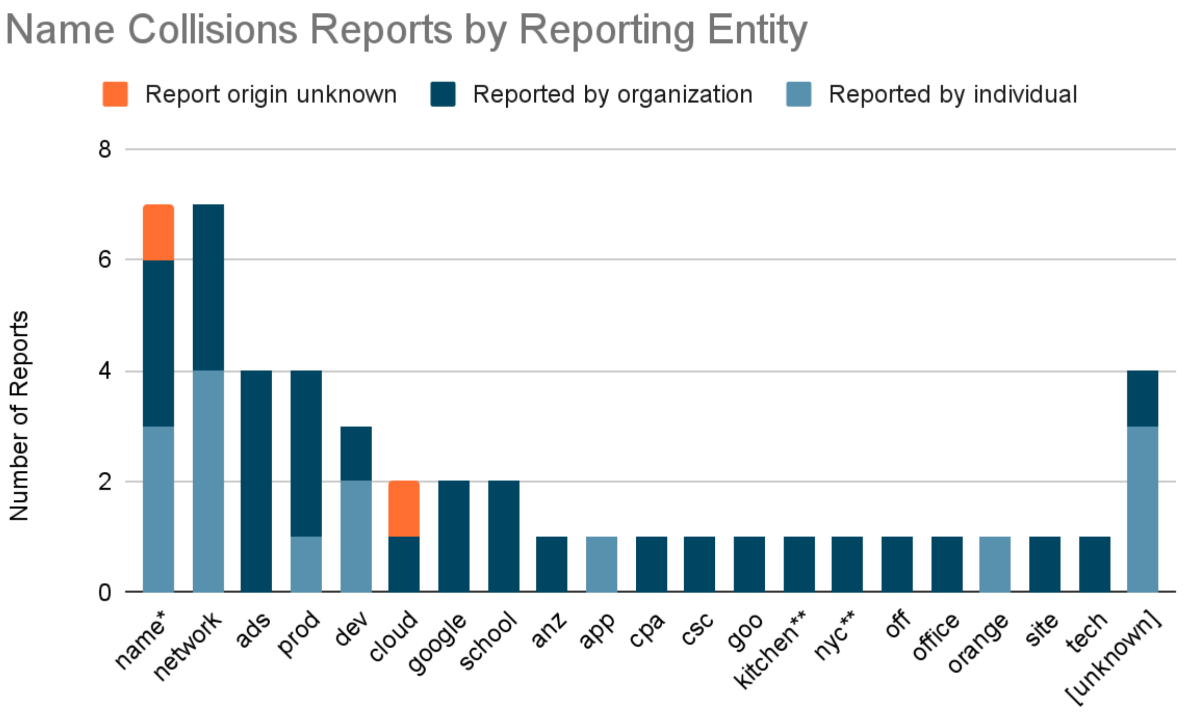
The following plot shows the distribution of reports by TLD, including those that were not delegated as part of the new TLD program (*) and those that were delegated prior to controlled interruption (**). For the 17 reported TLDs that were delegated after controlled interruption was introduced, each bar in the plot is composed of the numbers of reports received during and after the controlled interruption period for the TLD.



In most (74%) cases, report(s) were submitted during the controlled interruption period for the TLDs; in the remaining cases, the report was submitted after the controlled interruption period. For three TLDs, (`dev`, `app`, and `cpa`), *all* reports came *after* the controlled interruption period. It is unclear why reports for these three TLDs or for any others were submitted after the controlled interruption period. No other report details provided by the submitter shed any additional light.

Reporting Entity

Reports were categorized as having been submitted on behalf of an organization, submitted by an individual, or for which the origin was unknown. Considering only the 34 reports for TLDs delegated after the introduction of controlled interruption, the counts were 24 (71%) by organization, 9 (26%) by individual, and 1 (3%) unknown. The breakdown is shown in the following plot, which includes TLDs that were not delegated as part of the new TLD program (*) and those that were delegated prior to controlled interruption (**).


Name Collisions Reports by Reporting Entity

Two standouts are `ads` and `school`, for which reports were made exclusively by organizations. `ads` (as well as `local` and `intern`) is reportedly (as indicated in one report, but not independently verified) used in books and trainings for creating Microsoft Active Directory domains. Other reports indicated that `office`, `off`, `school`, and `site` are used by organizations for Active Directory services. `school` is reportedly used by some school districts as a private DNS namespace and—at least in some cases—for Active Directory, as mentioned previously.

Impact

In addition to the quantitative analysis associated with the affected TLDs and their reporting organizations, we now use additional report details to add a qualitative analysis. We consider only the 34 reports associated with TLDs delegated after the introduction of controlled interruption.

We first categorize impact based on the self-reported description and size of organization, if reported. We group incidents into four categories based on what we could infer from the content of these fields:

- *severe*. A large number of users were affected, network access as a whole was affected, and/or the submitter described the impact as severe.
- *significant*. The number of affected users or systems was more moderate, and/or only specific network applications were impacted.
- *small-scale.* The number of affected users or systems is small, and/or impacts seem nominal.

- *unknown*. There is insufficient data in the report to justify assignment to one of the other categories.

In the following table, we list the count for each category as well as sample comments from each report that led us to categorize them accordingly (except for unknown, for which details were too few to categorize otherwise):

Category	Count	Descriptions
Severe	7	“more 30,000 employees in over 7 countries and these employees interact with one another and with the organization via an internal network…. employees had trouble accessing their internal network.” “Network down, no internet access” “this is causing all of our staff laptops to crash when off of our network… this is causing severe problems” “All clients are having problem and freeze during usage.” “This is affecting all users in the organisation at various times” “1400 servers in 800 schools” “The scale of the impact is fairly critical. All VPN tunneling to our network cannot resolve DNS…. it is affecting all of our external users needing to resolve anything internal via DNS name. 300 users affected. All systems that reside outside of the office…”
Significant	10	“CRM, MAIL and other Services provided by our Company do not work correctly” “Unable to send mail” “150 users” “No network shares access.” “Do not operate normally computers are connected to a domain controller” “VPN sessions with split tunnelling do not work as the DNS lookup fails.” “If our applications are started before the corporate VPN connection is up… we cannot use the app's anymore” “Unable to resolve internal Hostnames” “some Clients… not correct working with the DNS Suffix Searchlist” “Users cant loggon to local domain”
Small-Scale	10	“Internet browsing issues from LAN” “can't access to some servers” “home network disruption” “Having trouble connecting to some network resources” “i cant use my sub domain… any longer”
Unknown	7	
Total	34	

While *some* impact *must* have been felt by all submitters—as evidenced by the mere existence of a report—our analysis shows that half (50%) were classified as either severe or significant.

Root Cause Identification

Clearly, all 34 reports were led to ICANN's name collisions report page to submit the report. Of the 34 reports, 8 (24%) specifically either mentioned "127.0.53.53" or referred to "controlled interruption" by name. It is unclear from the other reports whether the controlled interruption IP address itself contributed to finding the ICANN form, but we can say that at least one quarter observed 127.0.53.53.

Other Observations

We here record two significant trends that we observed in our analysis of the reports.

First, 8 (33%) of the 24 reports submitted by organizations mentioned "remote users" or "VPN" (Virtual Private Network). A VPN is typically used to connect the systems of these users to the corporate network. Once VPN-connected, the remote system typically uses the corporate DNS servers, but prior to connection, they must use a non-corporate (i.e., "public") DNS resolver. A common configuration for organizations using private DNS namespaces is for the corporate DNS resolvers to be configured to answer authoritatively for the private DNS namespace. This "works" when corporate systems *only* ever issue queries to the corporate DNS resolver—not to the public DNS. However, as evidenced by the submitted reports analyzed in this section, observed leakage of DNS queries for private DNS namespace (see Section ??), and responses to our survey (see Section ??), this is not always the case.

Second, of the 24 reports submitted by organizations, 8 (33%) explicitly mentioned Active Directory services. One additional report did not mention Active Directory, but the associated TLD was `ads`, so it might be inferred. Three (37%) of the reports mentioning Active Directory *also* mentioned VPN usage, i.e., that it was the combination of the two that caused the disruption. This shows that the impact of name collisions on systems using Active Directory are not isolated.

Leaked Suffix Identification

The queries in DNSDB-CI provide a look into the quantity and nature of controlled interruption queries being issued. This is enlightening because it corresponds to DNS queries leaked—whether intentionally or unintentionally—to the public DNS. These are queries which, prior to controlled interruption for the given TLD, would have resulted in an NXDOMAIN response from the root servers. Finding a meaningful way to systematically measure these queries is the next important step in our analysis.

Typical metrics for quantifying the DNS query activity associated with a given TLD include query count, IP address distribution, ASN distribution, second-level domain (SLD) distribution, and query name (qname) distribution. Unfortunately, of all these metrics, only one is feasible and useful: the query count— both per-qname and per-TLD. While IP address and origin ASN would be useful, neither is available with DNSDB. This is because DNSDB only provides a mapping of domain name to a resource and a query count associated with each mapping—no query source information. The diversity of SLDs and query names is only an effective measure inasmuch as there is additional context to understand how to categorize those SLD and qnames. For example, consider the qnames `foo1.bar.baz.com` and `foo2.bar.baz.com`. These are certainly distinct qnames and can be counted as such. But when considering the organizational diversity of these names, the question might be asked: do they originate from the same organization? This is difficult to know with only the qnames themselves, but if we had additional contextual data indicating that the DNS suffix (i.e., the right-most set of labels) `bar.baz.com` is common for a given organization, then that increases confidence that they do in fact originate from the same organization. Similarly, qnames `foo.bar1.baz.com` and `foo.bar2.baz.com` are clearly from the same SLD, but there is insufficient data in the names themselves to assert that they are from the same organization. For example the domains `state.ut.us` and `k12.ut.us` are delegated to two different entities, even if they have a common SLD.

Rather than using qnames or SLDs, we identify *DNS suffixes* to apply our query metrics (see section ??). This allows us to more effectively measure the nature and diversity of DNS queries because each query can be associated with a given network configuration setting that would be expected to be applied consistently to systems in the administering organization.

Our analysis applies three heuristic techniques to identify these DNS suffixes, given a set of queries: Chrome NXDOMAIN probing, WPAD lookups, and ISATAP preferred router lookups. In all three cases, we use the DNSDB-CI data set to provide the queries.

Suffix Identification via Chrome NXDOMAIN Probing

The first method of DNS suffix identification involves inferring Chrome NXDOMAIN probing from DNS queries observed in the DNSDB-CI data set. Any such activity would indicate Chrome browser usage, suggesting it originated from end-user application usage. Additionally it would identify the DNS suffix in use by the respective systems and users.

We note that queries associated with Chrome NXDOMAIN probing would not normally be found with DNSDB queries because, by definition, there is no mapping associated with NXDOMAIN responses. However, during the controlled interruption period for a TLD, *all* queries for qnames under that TLD result in an answer. Such is the case with the DNSDB-CI data set.

We now explain the procedure we employed to identify NXDOMAIN probing behavior. Chrome sends three DNS queries, all with the same DNS suffix, each with a randomly-generated first label, and all in rapid succession. Therefore, we look for DNS mappings (i.e., associated with

DNS queries) exhibiting that pattern. We use DNSDB's "first seen" timestamp to group mappings first observed at a given timestamp. We then considered all mappings observed at each timestamp, according to the following criteria:

- **First label.** Only mappings for which the first label of the domain name had a length of between 7 and 15 characters consisting of all alphabet letters were considered.
- **Query type.** Only mappings for which the query type was A (address) were considered.
- **Qname observed only once.** Because the first label of the qnames related to Chrome NXDOMAIN probing are randomly generated, it is probabilistically unlikely—though not impossible—that the same qname would be observed more than once. Thus, we only consider mappings for which the "first seen" timestamp equals the "last seen" timestamp, i.e., it was only observed once.
- **Qname group by suffix.** Mappings were grouped by common suffix—defined as everything after the first (i.e., left-most) label. Only suffix groups having a size that was a multiple of three were considered.
- **Qname group only seen once.** Only groups of qnames observed exactly once were considered because of the improbability of observing two groups of randomly-generated qnames that were exactly the same.

The list that resulted consisted of the suffixes (i.e., everything after the first label) for every qname group that met the criteria above. For example, suppose the following queries were observed:

First seen	**Last seen**	**Query (qname/type)**
1649687014	1649687014	`sujenbfd.foo.example.com/A`
1649687014	1649687014	`pwfiksd.foo.example.com/A`
1649687014	1649687014	`nmzuhes.foo.example.com/A`
1649687014	1649687017	`lkaubqq.foo.example.com/A`
1649687020	1649687020	`polkuhadev.bar.example.com/A`
1649687020	1649687020	`fvqiyjas.bar.example.com/A`
1649687020	1649687020	`hnsjmirc.baz.example.com/A`

This query data would result in the following DNS suffix: `foo.example.com`. Other potential DNS suffixes are not part of the resulting set because they do not meet all of the aforementioned criteria.

Despite the methodology matching

Suffix Identification Using WPAD and ISATAP DNS Queries

To identify suffixes using DNS queries related to WPAD and ISATAP, we identified all qnames with first label was “wpad” or “isatap”, respectively. The suffix list was built by extracting the suffix (i.e., everything after the first label) from every qname beginning with “wpad” or “isatap.”

Results

The total number of DNS suffixes identified in the DNSDB-CI data set was 2,761. Each identification method contributed to identifying these suffixes. The following table shows the *individual* contributions of each of the suffix identification methods—that is, how many of the suffixes were identified only because of the listed methodology was employed:

Methodology	**Suffixes Identified Exclusively by Method**
WPAD	360 (13%)
ISATAP	453 (16%)
Chrome	197 (7%)
Combined	2,761 (100%)

The ISATAP methodology was the single largest contributor, from which 16% of the suffixes were identified. The Chrome NXDOMAIN probing had the lowest individual contribution, yet without it, 7% of DNS suffixes would not have been identified.

While at least one suffix was found in 498 (56%) of the 885 new delegated TLDs, the distribution of suffixes across TLDs was such that most of the suffixes were concentrated within a relative few. The following table shows a per-TLD statistical breakdown of the suffixes, both overall and by individual identification method:

	Number of Suffixes per TLD			
	Median	**90th percentile**	**99th percentile**	**Max**
WPAD	0	3	37	223
ISATAP	0	3	40	240
Chrome	0	2	27	145
Combined	1	3	52	297

Thus, half of TLDs were associated with at most one suffix, and fewer than 10% of TLDs were associated with more than three suffixes. Particularly interesting is the high correlation between the number of DNS suffixes identified in newly delegated TLDs and their inclusion in reports submitted via ICANN’s Web form. The following table lists each reported TLD, in order of rank, along with the numbers of DNS suffixes identified in each. Only the 17 TLDs delegated after controlled interruption (August 2014) are included, as they are the only ones for which we have suffix data from the DNSDB-CI data set *because* of controlled interruption. Numbers that are underlined indicate value above the 90th percentile.

TLD	**ICANN Reports**	**DNS Suffixes Identified Using Method**			**Total DNS Suffixes Identified**
		Chrome	**WPAD**	**ISATAP**	
network*	7	60	86	115	134
ads*	4	139	233	234	247
prod*	4	32	64	66	71
dev*	3	62	100	98	113
cloud*	2	10	14	12	14
google**	2	1	6	3	3
school*	2	29	37	40	47
anz	1	0	2	0	2
app*	1	3	3	5	6
cpa*	1	2	6	3	4
csc	1	2	2	2	3
goo	1	0	1	1	1
off*	1	7	15	14	14
office*	1	145	216	240	264
orange*	1	3	5	4	5
site*	1	18	23	33	50
tech*	1	18	25	30	33

* All DNS suffix counts were in the 90th percentile.
** At least one DNS suffix count was in the 90th percentile—but not all counts were.

At least one DNS suffix was identified for every TLD for which problems were reported, and all reported TLDs except one (goo) had suffix counts greater than the median. In 13 (76%) of the 17 TLDs for which problems were reported, the number of DNS suffixes were in the 90th percentile. In only 3 (18%) of the 17 TLDs for which reports were submitted were all suffix counts below the 90th percentile. Further, the 4 (24%) TLDs with the most reports (i.e., the four highest ranking) had suffix counts within 99th percentile.

The trends here are clear. There are disproportionately high counts of DNS suffixes amongst the 17 reported TLDs, with 76% having DNS suffix counts in the 90th percentile. The trend clearly suggests that reports for a given TLD are more prevalent where the DNS suffix count is higher.

Validation of WPAD Identification Methodology

As mentioned previously (TODO), there was some question about false positives produced when using the WPAD identification methodology. Specifically, there was some concern that "ancestor" names of a legitimate DNS suffix might be falsely identified as DNS suffixes because of the iteration performed by WPAD clients. We evaluated our results to look for evidence of such behaviors.

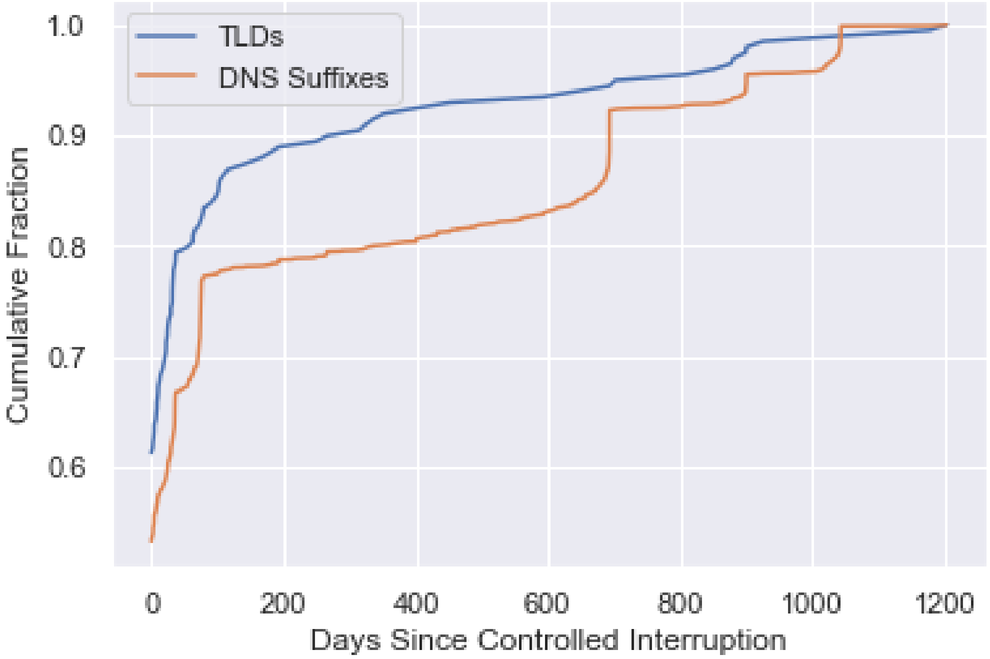
Of the DNS suffixes using the WPAD identification methodology, 1,728 suffixes were composed of two or more labels. For only those cases, only 153 (8.9%) was the "parent" DNS name also identified as a suffix using the WPAD methodology. In 91 (59%) of those cases, the parent name was identified independently as a DNS suffix using one of the other methodologies. Thus, in only 62 (3.6%) of cases was a parent name identified *exclusively* by our WPAD methodology as a DNS suffix. It is possible that every one of these "parent" suffixes is a legitimate DNS suffix, but even if not, the low percentage shows that this is not a pervasive behavior.

Controlled Interruption Analysis

We use the DNSDB-PostCI data to learn more about the use of controlled interruption and the use of the observed DNS suffixes identified as being in conflict with new TLDs being delegated. By considering only DNS suffixes that had two or more labels (see also section ??), we reduced the numbers of DNS suffixes and TLDs to 2,300 and 200, respectively, and looked at the mappings observed since the first 100 days of delegation for each DNS suffix. Note that this filtered set of DNS suffixes included 16 (94%) of the 17 TLDs reported to ICANN; only the `goo` TLD (associated with a single ICANN report) was excluded.

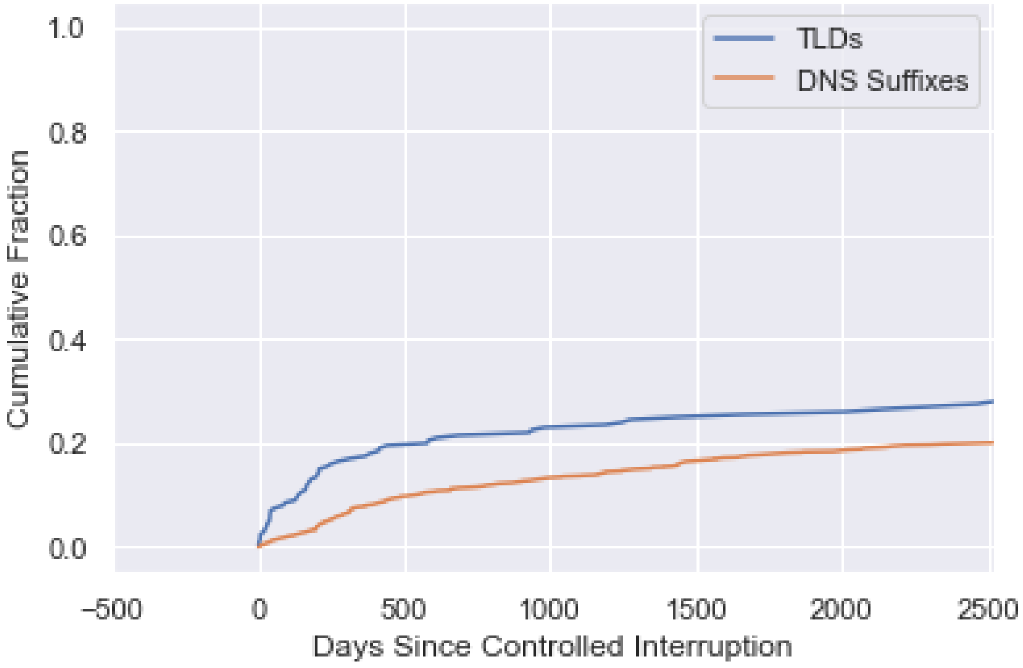
As mentioned previously (Section ??), the IP address 127.0.53.53 is returned for all names under a TLD during the first 100 days of its delegation, i.e., the controlled interruption period. By analyzing the mappings in DNSDB-PostCI, we were able to determine how long controlled interruption was observed for each TLD and at what point non-controlled interruption addresses (i.e., other than 127.0.53.53) were observed in relation to the controlled interruption period.

The following plot shows the cumulative distribution of the number of days after the controlled interruption period for which the controlled interruption address was observed—on a per-TLD basis and a per-suffix basis:



For about 53% of DNS suffixes and 62% of TLDs, the controlled interruption address was not observed after the controlled interruption period, i.e., the first 100 days of delegation. However, the controlled interruption IP address was observed for a year or more after the controlled operation period for about 10% of TLDs and for 20% of DNS suffixes.

While a glimpse of how long controlled interruption was maintained beyond the prescribed period, perhaps more interesting and useful is an understanding of how soon after the controlled interruption period non-controlled interruption addresses were introduced for suffixes known to be used in conjunction with private DNS namespaces. The following plot shows the cumulative distribution of days since controlled interruption representing those mappings:



For about 72% of TLDs and 80% of DNS suffixes, no mappings were observed for known DNS suffixes. However, for the remaining 28% and 20% of TLDs and suffixes, respectively, non-controlled interruption mappings were observed at some point after the controlled interruption period ended. In both cases, those mappings were observed immediately after; for 10% of suffixes and 20% of TLDs mappings were observed within 500 days (about 16 months).

The presence of non-controlled interruption does not pose an immediate threat in and of itself; it all depends on the existence of a mapping for a qname within a DNS suffix and, of course, the nature of the application or service relying on the resolution. However, it does indicate the *potential* for third-party and interception of traffic, whether intentionally or inadvertently. While we have not carried out a general search of qname mappings, we did search for two prominent qname patterns, which, if present, could have a significant impact on systems relying on the non-resolution of certain DNS qnames used for private use: `wpad` and `isatap` (see sections ?? and ??). Fortunately, we found no mappings for such qnames in the DNSDB-Post-CI data.

Root Server Query Analysis

The DNS suffixes identified in Section ?? provide a unit of measurement for quantifying the usage of newly-delegated TLDs, prior to and after their delegation, and to identify organizations from which their associated queries originated. In this section we describe our measurement methodology.

We used the DITL data from 2014 through 2021 (see section XX) to observe queries at the root servers related to the DNS suffixes associated with leaked DNS queries, i.e., those identified previously. Extracting query information from the DNS root servers requires resources related to both computation and storage. For this reason, we reduced the computational resources required by limiting the suffixes against which we compared DITL queries in two ways. First, we reduced the suffixes by eliminating those that were themselves TLDs. For example, both `office` and `mercury.office` were identified as suffixes through one or more of the identification techniques. While `mercury.office` remained in our data set, `office` was eliminated because it was a TLD. A DNS suffix with only a single label is typically too generic to help us characterize queries with that suffix.

Second, we further limited our analysis to suffixes with TLDs meeting one or more of the following criteria:

- The number of DNS suffixes identified from ISATAP-related queries was at least one;
- The number of DNS suffixes identified from WPAD-related queries was at least one; or
- The number of total DNS suffixes identified as at least two.

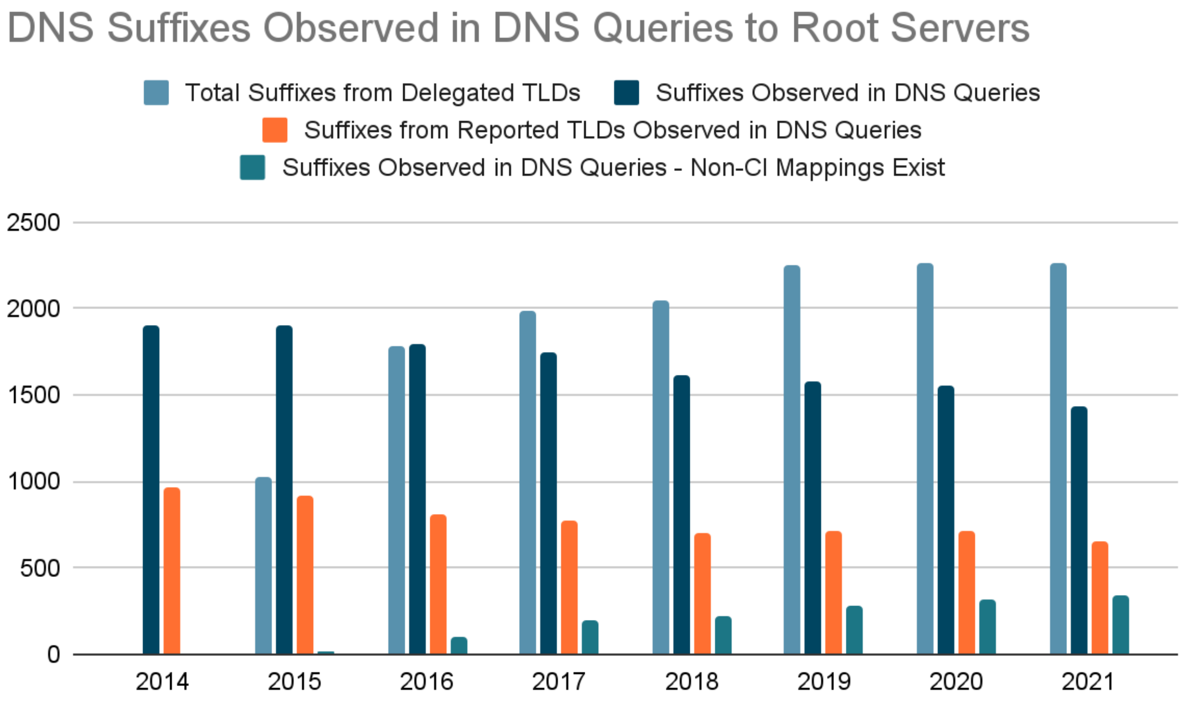
This effectively eliminated DNS suffixes for TLDs that were *only* part of the data set because of a single suffix identified with our Chrome NXDOMAIN probing technique. While all three of our suffix identification techniques were merely heuristics, Chrome NXDOMAIN probing was the most susceptible to false positives. This filter eliminated some of the weaker contributors in the data set.

	Suffixes	**TLDs**
All DNS Suffixes	2,761	498
DNS Suffixes - no TLDs	2,300	200
DNS Suffixes - no TLDs and further filtered **(TLD has at least one WPAD suffix, one ISATAP suffix, or more than 1 total suffix)**	2,266	166

Note that this filtered set of DNS suffixes included 16 (94%) of the 17 TLDs that were the subject of reports submitted to ICANN via their Web submission form (see section ??). The only TLD that was excluded was `goo`.

Having our updated DNS suffix list in hand, we utilized a two-step process to actually extract the DNS queries from the DITL: 1) we filtered all DITL queries, keeping only those with a query name under one of the newly-delegated TLDs; then 2) we tested each of the resulting queries to see if the query name was under one of the 2,266 DNS suffixes we identified previously.

We first consider the number of DNS suffixes observed in root queries during each DITL collection period between 2014 and 2021. The following plot shows: 1) the total number of DNS suffixes for which their TLD was delegated during the time of the DITL collection for the corresponding year (i.e., all 2,266 were delegated by the time of the 2021 DITL collection); 2) the total number of DNS suffixes for which DNS queries were observed at the root servers, out of the 2,266 total suffixes; 3) the subset of observed DNS suffixes that were the subject of ICANN reports (see section ??); and 4) The number of DNS suffixes for which DNS queries were observed and for which non-CI mappings (i.e., other than 127.0.53.53) were identified after the CI period for the respective TLD (i.e., after the first 100 days).

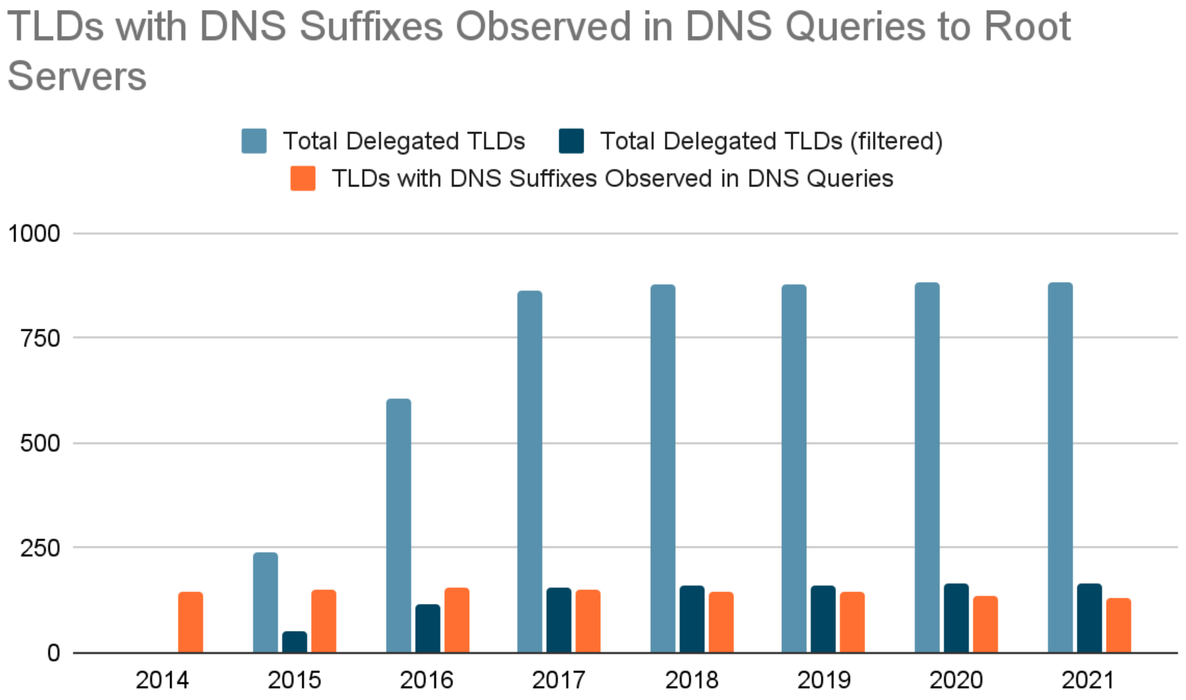


While over 1,900 (84%) of the 2,266 DNS suffixes were observed as early as 2014, the number of suffixes observed in DNS queries has consistently decreased over time, as new TLDs have been delegated, such that in 2021 1,434 (63%) suffixes were observed. Nearly half of those DNS suffixes are associated with the reported TLDs, specifically between a low of 43% (2018) and a high of 51% (2014). This disproportionately high contribution of observed DNS suffixes again emphasizes the significance of the name collisions reports submitted to ICANN.

We note that *all* of these suffixes were observed during the controlled interruption period for their respective TLDs and have thus been associated with leakage of “private” DNS queries colliding with public DNS namespace. However, we cannot know from these query observations alone whether the queries at the root were associated with previous, private use of the TLD (i.e., prior to its delegation) or use of the TLD in connection with its delegation. The latter is certainly the case in 2014 because none of the new TLDs or their suffixes were delegated by the time of the 2014 DITL collection, but for 2015 and beyond, it is not known. See section ?? for more.

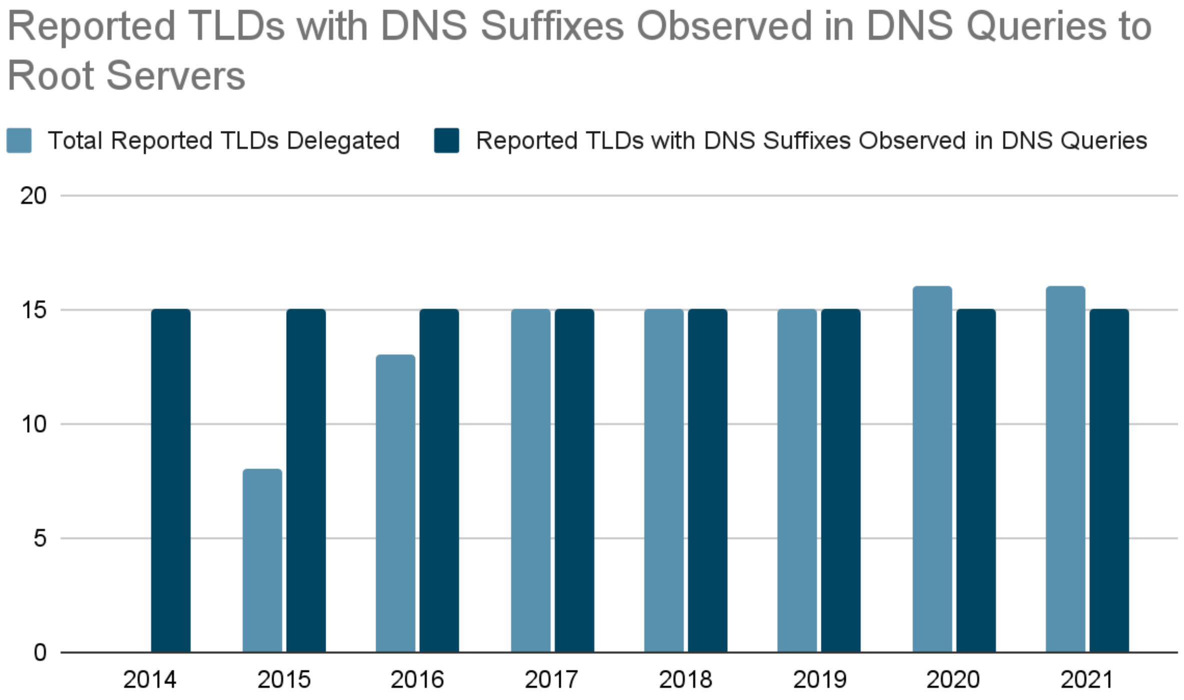
Between 2015 and 2021, there is a steadily increasing number of DNS suffixes observed in query data for which non-CI mappings exist (see section ??). In 2021, 336 (23% of filtered, 15% of all) DNS suffixes had a non-CI mapping. As mentioned, it is difficult to tell with current data whether the queries associated with these suffixes were in connection with private use or not, but it does raise some concerns.

We now consider the same data, but with respect to TLD. The following plot shows: 1) the total number of TLDs delegated during the time of the DITL collection for the corresponding year (i.e., a total of 885 delegated TLDs by the time of the 2021 DITL collection); 2) the total number of *filtered* TLDs delegated at the time of DITL data collection (i.e., a total of 166 TLDs by the time of the 2021 DITL collection; and 3) the total number of TLDs having DNS suffixes for which DNS queries were observed at the root servers, out of the 166 filtered TLDs. In other words, this plot shows the number of TLDs experiencing some sort of name collision behavior over time.



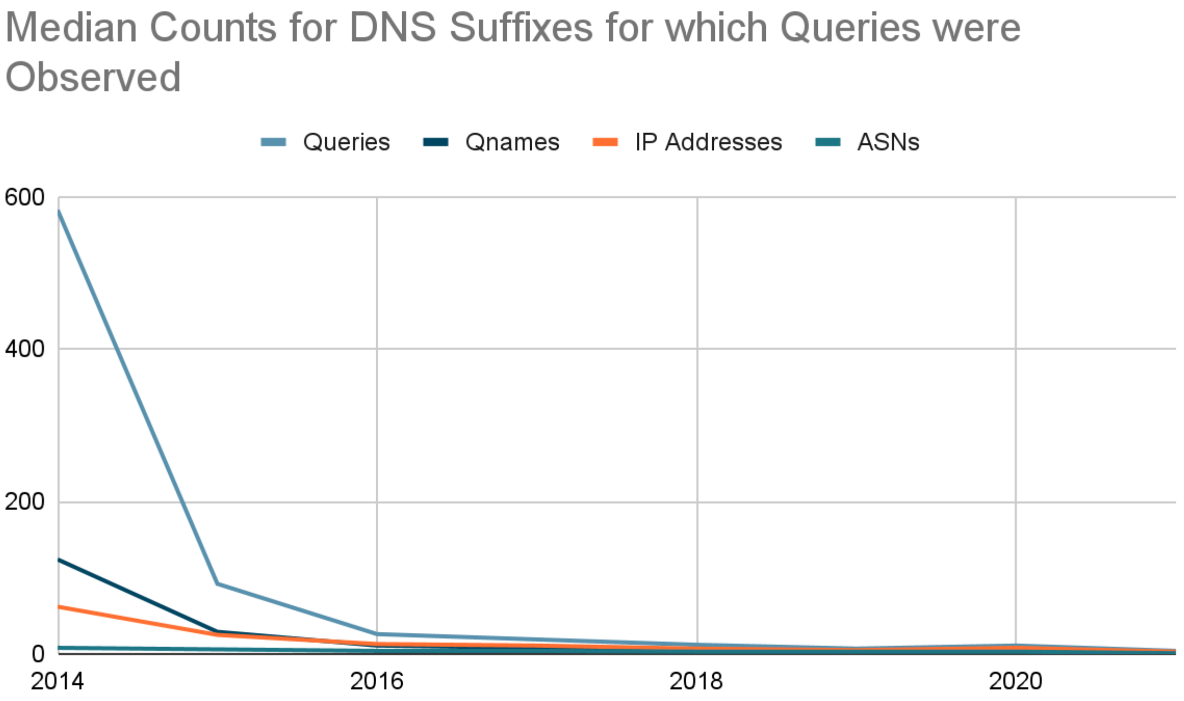
The number of TLDs experiencing name collisions, by observation, has remained relatively steady from 2014, when queries for DNS suffixes associated with 146 TLDs (88% of filtered, 16% of all TLDs) were observed, through 2021, when 133 TLDS exhibited name collision behavior (80% of filtered, 15% of all TLDs). The peak was in 2016 when 154 TLDs (93% of filtered, 17% of all TLDs) exhibited name collision behavior.

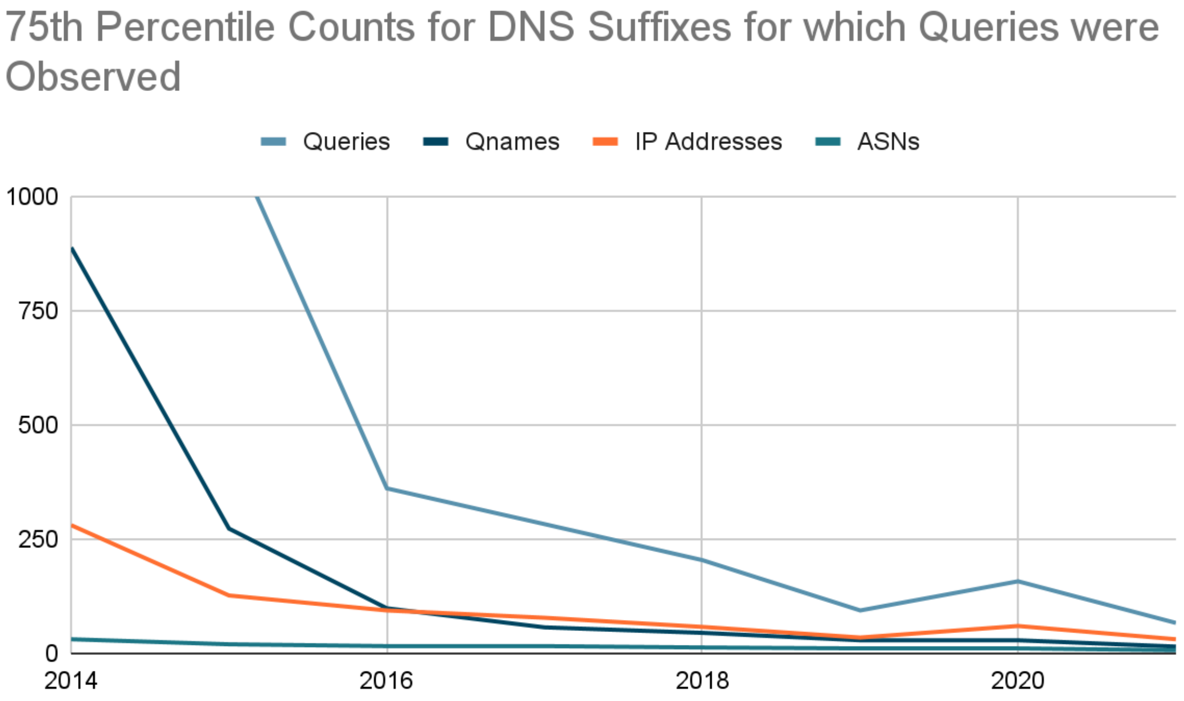
When we consider only the 16 TLDs that were the subject of reports and part of the filtered set of DNS suffixes, the following plot is the result:



This shows that in every DITL collection between 2014 and 2021, queries for DNS suffixes within 15 (94%) of the 16 reported TLDs, after filtering, were consistently observed. Only t TLD `google` was not observed. While the general trend was mostly consistent, this trend was completely consistent.

We now consider several other metrics to help us quantify name collision behavior between 2014 and 2021. Specifically, for DNS suffixes experiencing queries each year, we consider the number of queries, unique qnames, querying IP addresses, and origin ASes of queries. The median and 75th percentile values are shown in the following two plots:





In both plots, a clear trend of decreasing per-suffix usage metrics is evident. However, the cause of this trend is unknown. One possible cause might be actual administrative changes eliminating the use of those suffixes in configurations, possibly because of the effects of controlled interruption. However, it could also be due to reduced DNS query data at the root servers associated with local root deployments (RFC 8806) or qname minimization (RFC 7816). The former, which was first published in 2015 and updated in 2020, provides guidance for serving a copy of the root zone on a recursive resolver. This keeps the resolver from having to issue any queries the root servers because it has all the answers it needs locally. It thus achieves benefits of both privacy and performance. With qname minimization, a recursive resolver only reveals the necessary parts of the name it is attempting to resolve in the queries it issues to authoritative DNS servers. For example, when a resolver is resolving `www.example.com`, it might have historically sent the entire name, `www.example.com`, to a root server. However, a qname-minimizing resolver sends only `com` (or `_.com`) knowing that that is all that is needed to elicit the proper referral response to the `com` servers. Recent studies suggest that qname minimization affects XX% of Internet resolvers and XX% of queries as of YYYY. However, to date, there are no research studies to provide insight into the prevalence of local root deployment.

To gain additional insight into the causes of the behavior we observed, we supplement our quantitative measurements with a qualitative study, which we discuss in the next section.

Name Collisions Survey

To better understand the metrics we presented in the previous section, we conducted a survey to solicit experiences related to name collisions. The survey was given to two different target audiences: a general audience of network operators and a targeted audience consisting of organizations presumably affected by name collisions related to the delegation of new TLDs.

Survey Content

The questions were common to both surveys, with some slight variants in wording. They solicited the following information:

- What DNS suffixes under newly delegated TLDs are in use by organizations.
- Which newly delegated TLDs are associated with DNS suffixes in use.
- What DNS configuration is being used in the organization in connection with suffix use.
- Whether or not problems were experienced with the use of the DNS suffixes since the delegation of the TLD.
- What the effects of suffixes were, in terms of time to detection, number of users affected, and time to resolution.
- What was the role of controlled interruption IP address (127.0.53.53) in diagnosing the problem.

The complete set of survey questions for the general and targeted audiences are found in Appendixes A and B, respectively.

Survey Recipients

The general version of the survey was sent to the North American Network Operators Group (NANOG) mailing list on March 29, 2022, with a reminder email sent on April 4, 2022.

The recipients for the targeted version of the survey consisted of network administrators for which the autonomous system (AS) description matched DNS suffixes corresponding to queries originating from that AS number (ASN). We created this list using the following methodology:

- Create suffix-ASN mappings from queries observed at root servers, based on DITL data (section ??).
- Filter suffix-ASN mappings to include only suffixes for which at least 10 unique qnames (implies at least 10 queries) were observed for the suffix for any collection year. This filter was used to establish additional confidence in the sample set of suffixes that would be used for targeted reach-out.
- Further filter suffix-ASN mappings to include only ASNs that included a single suffix. This filter is applied to exclude ASNs that likely provide a DNS resolver service for other organizations.
- For each suffix-ASN mapping, perform a WHOIS lookup of the ASN, and compare the organization information provided by WHOIS with the DNS suffix itself (typically the left-most label). Include only mappings for which a positive match was made.

This process resulted in a list of 28 mappings in 18 TLDs for which we could associate ASN technical contact information. These included 7 (44%) of the set of 16 reported TLDs (after filtering). However, there was no selection bias based directly on report TLDs; we selected all mappings from the sample for which we were able to positively identify a match between DNS suffix and ASN.

The targeted messages sent to ASN contacts contained not only a link to the survey, but also the DNS suffix associated with the mapping—that is, the one for which DNS queries were observed as having originated from the ASN.

One known limitation of our methodology is that the mappings consist of DNS suffixes that match the ASN descriptions; however, one of the observations made in section ?? is that a significant contributor to name collisions is systems querying the public DNS from *outside* their corporate network (in which the DNS resolvers might be configured to answer authoritatively). Thus, the targeted survey results have some bias related to the symptoms and possibly the network configuration causing the issues. The targeted surveys might also represent a community with a private query leakage caused by something different than the remote user/VPN configuration noticed in section ??. However, as will be noted, this bias has little impact on our findings because the response rate was so low.

General Survey Results

The survey sent to the NANOG mailing list generated 31 responses. Of those 31, 21 (68%) indicated that their organization did not employ any DNS suffixes that were associated with newly delegated TLDs. We focus the remainder of this analysis on the 10 (32%) respondents that indicated that they *did* use DNS suffixes under new TLDs.

TLDs Used

The following tables lists the TLDs associated with survey responses, representing DNS suffixes in use by organizations:

Delegated Before Controlled Interruption	**Delegated After Controlled Interruption**	**Not Delegated**
audio	dev*	corp
foo	group	example
media	llc	internal
pro	network*	test
	office*	
	tech*	

* Included in name collisions reports submitted to ICANN.

Most pertinent to this root cause analysis are the TLDs in the middle column, which represent the TLDs that have been delegated since controlled interruption (i.e., since August 2014). Four of those (marked with *) were also the subject of reports submitted to ICANN via their Web form.

Technical Issues Experienced

Of the 10 reports in which DNS suffix use was indicated, 7 (70%) reported experiencing technical problems after delegation of the TLDs. We focus our analysis on just those 7 reports for the remainder of this section.

DNS Resolver Configuration

In three (43%) of the cases experiencing technical issues, the response indicated that the organization's configuration was such that the DNS resolvers were configured to answer authoritatively for the DNS suffixes in question; in two (29%) cases, that was *not* the configuration. Two respondents did not know details related to this configuration. There seems to be no strong correlation between the DNS resolver configuration and the presence of technical issues with the DNS suffix. Across the 10 responses confirming use of DNS suffixes within newly delegated TLDs and the 1 response confirming use from the targeted survey (**), we saw the following combinations:

DNS Resolver Authoritative	**Issues Experienced**	**Count**

No	No	2
No	Yes	2
Yes	Yes	2
Yes*	Yes	1
Yes*	No	1
Yes	No	1**

* Resolvers were changed to answer authoritatively at some point.

** Included from the targeted survey response.

Discovery, Impact, and Resolution

Three (43%) organizations discovered the problems within days of the delegation; one (14%) within weeks of the delegation; and three (43%) within months of the delegation. In terms of impact, three (43%) reported that only a few systems were affected, but two (29%) reported that many were affected, and two (29%) reported that nearly all systems were affected. Two (29%) reported that they were able to resolve the issue within days or weeks of its discovery. However, two (29%) reported that it took years to resolve, and two (29%) reported that it has not yet been resolved.

Root Cause Identification

With respect to the identifying the root cause of the problem, five (71%) respondents indicated that they knew the problems were related to the delegation of new TLDs before the problem was resolved, and two (29%) only discovered that the problems were related to delegation of new TLDs after the problem was resolved. In only one (14%) case was the controlled interruption IP address, 127.0.53.53, observed and helpful in leading the organization to ICANN and the delegation of the new TLD. One (14%) respondent reported that 127.0.53.53 was observed, but its meaning was unclear and was not helpful in identifying the problem. In the five (71%) remaining cases, 127.0.53.53 was not observed at all.

Other Observations

Some of the free-form comments received from respondents shed additional light on the experiences of those who were impacted by new TLD delegations.

One respondent indicated that their DNS resolvers were not configured as authoritative for their DNS suffix, but rather for the entire TLD (`dev`). The problems then came when `dev` was delegated. In this specific case, they reported that the problem was discovered within days of its delegation, affected “many” users or systems of an organization with fewer than 1,000 systems, and took weeks to fix. The fix involved changing the DNS suffix they were using internally (e.g., as opposed to changing the way their DNS resolvers were configured). In this case, 127.0.53.53 was not observed.

Another respondent commented:

> "This was very expensive and disruptive. In addition, employees cannot reach websites in the network domain."

This response indicated that "nearly all" systems or users were affected by the change, in an organization consisting of between 1,000 and 10,000 systems. Although the problem was discovered within days of the delegation, it reportedly took years to fix. In this case, 127.0.53.53 was observed, but its meaning was unclear or unhelpful in identifying the problem.

Targeted Survey Results

Of the 28 targeted surveys, two recipients (7%) filled out the survey. Of those, only one recipient confirmed use of the suffix provided in the email message; the other was symptomatic of false positive match between DNS suffix and ASN.

The admin that confirmed usage of the provided DNS suffix provided the following information with regard to its use:

- The suffix is associated with the `win` TLD.
- Use of the DNS suffix predated the delegation of the TLD, and the DNS suffix continues to be used by the organization.
- The organization's DNS resolvers are configured to be authoritative for the DNS suffix, such that queries within those suffixes, when issued to their resolvers, are presumably not leaked to the public Internet.
- No known technical issues were experienced with the suffix after the delegation of its TLD.

Discussion

This work attempts to analyze several data sources consisting of mostly passive traffic data and couple that analysis with qualitative data from both a targeted and a general survey. We report here some of the key findings from the analysis, impact inferred from both quantitative and qualitative measurements, known and suspected limitations of this analysis, and proposed future work.

Findings

Private use of DNS suffixes is widespread. It is clear from the data that private use of DNS suffixes is not isolated. It is exhibited by over half of newly delegated TLDs, even a few TLDs are responsible for more usage than others. **Evidences.** Over half of the 885 TLDs delegated since August 2014 are being used as part of at least one configured DNS suffix for organizations, according to our measurements. Yet the use of DNS suffixes is not uniformly distributed across affected TLDs. Rather, 90% of TLDs are associated with three or fewer private-use DNS suffixes, but 1% have more than 52, reaching upwards of 297 (maximum).

Name collision reports correlate strongly with measured data. The TLDs appearing in name collision reports submitted to ICANN via their Web form rank disproportionately high in terms of the number of identified suffixes and DNS queries observed at the root servers. This bolsters the concerns associated with the reports and also indicates that there are likely others that experienced problems but did not submit reports. **Evidences.** About *two thirds (66%)* of reported TLDs were in the *90th percentile* of all TLDs for which DNS suffixes were identified, in terms of DNS suffix count. Additionally, TLDs associated with reports accounted for around half (between 43% and 51%) of the identified DNS suffixes that were observed in queries to the root servers, despite them comprising only 10% of the TLDs that were being watched for in the root server query data (i.e., the filtered set). Finally, while the observation rate of the *entire* filtered subset of TLDs ranged from 84% (2014) to 63% (2016), the fraction of *reported TLDs* for which DNS suffixes were observed in queries to the root servers was consistently 97%.

Usage of private DNS suffixes is decreasing over time. Various metrics related to DNS queries for DNS suffixes presumed to be used privately were measured over time and shown to be consistently decreasing since 2014. The reasons are unclear, but two considerations are 1) decreased DNS suffix usage and 2) reduced visibility at the root zones. **Evidences.** Both the median and 75th percentile counts of individual DNS queries, unique query names, querying IP addresses, and origin ASNs decreased sharply between 2014 and 2015, and have decreased more gradually since then. We have some anecdotal evidence of configuration changes from survey respondents, which supports the decrease. We also reference outside studies that show some uptake of qname minimization, which reduces the query context available at root servers (see section ??).

Controlled interruption is effective at disruption but not at root cause identification. Controlled interruption has shown to be good at disruption, but not at helping affected users identify the cause of the problem—at least not in the way that was intended. **Evidences.** Over two-thirds (71%) of survey respondents that experienced technical issues with a DNS suffix indicated that they knew that the issues were related to TLD delegation before the problem was resolved, and over half (57%) reported discovering the problem within days or weeks of the TLD's delegation. However, in 71% of cases, the controlled interruption IP address was not even observed, and only in one case (14%) was it shown to be helpful.

Configuring DNS resolvers as authoritative for DNS suffixes is not a panacea. DNS resolvers that respond authoritatively for private DNS suffixes do not prevent query leakage to the public DNS or name collision problems. **Evidences.** We have one confirmed account of DNS suffix usage where the queries were leaked to the public DNS: the targeted survey respondent confirmed usage of the DNS suffix, and we observed the queries within that suffix in the DITL query data. Additionally, the survey responses show no clear correlation between DNS resolvers thus configured and technical problems related to name collisions. In contrast, they show all combinations of issues experienced and resolver authoritative configuration. Further, 8 (33%) of the 24 ICANN reports submitted by organizations explicitly mentioned remote users or VPN usage.

The impact of TLD delegation ranged from no impact to severe impact. The only data we have quantifying impact related to delegation of new TLDs is from the name collision reports and the survey responses. With the limited responses we received, it is hard to generalize impact. However, what we *can* say from the data is that: 1) there is a range of impact reported, from no impact to major impact; and 2) there was evidence of both severe and significant impact amongst affected parties. **Evidences.** On one side of the spectrum, the one targeted survey respondent that confirmed DNS suffix usage indicated no technical issues. Seven respondents of the general survey indicated that they had experienced technical issues, with one describing it as “expensive and disruptive,” impacting almost all users or systems of an organization with between 1,000 and 10,000 systems. The remaining survey responses reported impact somewhere between no impact and extensive impact, based on both number of systems affected and total number of systems. In the name collisions reports, half (17 or 50%) of the reports imply severe or significant impact to the reporting entities.

Proposed Future Work

This work has provided many insights into the impact of the delegation of new TLDs since 2014. However, it also leaves many unanswered questions—along with some paths to answer them. Some of the trends in the measured data are clear: private DNS suffix usage appears to be declining; and the reports submitted to ICANN are supported by the measured data. However, the amount of qualitative survey data is far from adequate. It provides enough of a picture to see that experience has varied widely, ranging from no impact to high impact. Yet it is insufficient to complement and interpret the measurement data.

To fill the knowledge gap on the experiences of organizations, we propose additional work, targeting *analysis* and *reach-out* related to the suffix-ASN mappings. The goal in both of these is to better understand how DNS suffixes are being used and to further our understanding of organizational impact with TLD delegation. In performing the manual inspection and alignment of identified DNS suffixes and ASNs for a *small* sample, we gained experience and insight into the effort that might be applied to carry out the same work, more efficiently and effectively on a large sample. The key observation is that there are a variety of different suffix-ASN mappings, which are suffix-dependent, ASN-dependent, and network configuration dependent. We provide several examples below:

1. **Even statically configured systems are mobile.** While DNS suffixes are applied by an organization to its systems, some of those systems are mobile. Thus, even when a DNS suffix can be associated with a given organization and its ASN, queries for that suffix will appear from other ASNs, as mobile systems travel.
2. **DNS queries might never leak from their origin ASN.** Because of corporate DNS configurations in which DNS resolvers answer authoritatively to queries in private namespace, the leakage associated with the configuration of one ASN might *only* appear to originate from other ASNs.
3. **Many ASNs are ISPs.** These exhibit the characteristics that 1) they are more ephemeral in terms of suffixes observed; and 2) there are potentially larger numbers of DNS suffixes mapped to ISP ASNs because of mobile systems. These can be identified by name (e.g., “comcast”, “cox”, or “sprint”), but also by keyword (e.g., “mobile”, “wireless”, “telecom”, “cable”, or “broadband”).
4. **Generic suffixes are in use.** Generic DNS suffixes like `local.site` and `modem.local` are, by their very nature, not specific to any organization. Thus, the organization which is using it in its configuration is more difficult to identify.
5. **Regional subdomain suffixes are in use.** Some organizations have deployed suffixes globally, with region-specific subdomains. For example `corp.sap`, `homeaway.live`, `hsbc`, with labels like the following prepended: `emea`, `mos`, `de`, `aus1`.
6. **Some TLDs are commonly used for Active Directory services.** This includes `school`, `ads`, `site`, `prod`, and possibly others. And some books and trainings for Microsoft Active Directory direct administrators to use a private suffix, including some of the aforementioned TLDs.

We believe that using knowledge gained in this analysis, including the findings noted above, a more automated workflow could be developed to better match DNS suffixes to their origin organization. It is our hope that this will both enrich our understanding of the use of private DNS suffixes, create more opportunity for reach-out, and ultimately better understand past and future impact of delegation of new TLDs.

Appendix A - General Name Collisions Survey

Appendix B - Targeted Name Collisions Survey

Appendix C - General Email Sent to NANOG Subscribers

Dear colleagues,

tl;dr: Please take our survey on DNS suffix usage here: https://forms.gle/ntvsn6eqzYH9YcTN6

The Internet Corporation for Assigned Names and Numbers (ICANN) is researching the technical impact of delegating new generic top-level domains (gTLDs). This research is part of the Name Collision Analysis Project (NCAP). More information about NCAP can be found at https://community.icann.org/display/NCAP.

Since 2013 hundreds of new gTLDs have been introduced into the public DNS (https://newgtlds.icann.org/en/program-status/delegated-strings). In some cases those gTLDs might have been used as part of a DNS suffix by one or more organizations around the Internet, prior to their introduction. (By “DNS suffix” we mean a domain name used in the DNS resolver search list of a device, e.g., the “domain” and “search” entries in /etc/resolv.conf on UNIX/Linux, “Search Domains” in the macOS DNS configuration pane, and “DNS suffix search list” on Windows.) As a result, the behavior of systems or devices in these organizations might have changed because of a “name collision”. A name collision occurs when a name used in one context (in the organization's network) is interpreted in another context (in this case, in the public DNS after the corresponding gTLD went live).

We are researching the causes and impact of name collisions. We are seeking qualitative data based on experiences of those organizations potentially affected. We expect that this additional data will greatly enhance our understanding of name collisions that resulted from adding new gTLDs.

If you suspect that your organization has been impacted by the delegation of any new gTLDs, we invite you to please fill out the following brief survey regarding your experience. We would be grateful for your input!

https://forms.gle/ntvsn6eqzYH9YcTN6

Your responses will remain anonymous, and any personal information will be discarded after the research has concluded.

If you have any questions, please reply to this email.

Thank you for your help!

Sincerely,

Casey Deccio
ICANN Name Collisions Analysis Project

Appendix D - Targeted Email Sent to AS Contacts

Dear network administrator,

The Internet Corporation for Assigned Names and Numbers (ICANN) is researching the technical impact of delegating new generic top-level domains (gTLDs). This research is part of the Name Collision Analysis Project (NCAP). More information about NCAP can be found at https://community.icann.org/display/NCAP.

Based on our research, we believe systems or devices in your organization might have been using the DNS suffix "«DNSSuffix»" when the top-level domain "«gTLD»" was added to the DNS root zone on «Date». (By "DNS suffix" we mean a domain name used in the DNS resolver search list of a device, e.g., the "domain" and "search" entries in /etc/resolv.conf on UNIX/Linux, "Search Domains" in the macOS DNS configuration pane, and "DNS suffix search list" on Windows.) We inferred possible use of this DNS suffix by analyzing several years of DNS queries captured at the DNS root servers as part of the annual Day In the Life (DITL) collection (https://www.dns-oarc.net/oarc/data/ditl). We used publicly available WHOIS information for your autonomous system to find your contact information and send this email.

After the TLD «gTLD» went live, the behavior of systems or devices in your organization might have changed because of a "name collision". A name collision occurs when a name used in one context (in this case, inside your organization) is interpreted in another context (in this case, in the public DNS after «gTLD» went live).

We are researching the causes and impact of name collisions. We are seeking qualitative data based on experiences of those organizations potentially affected. We expect that this additional data will greatly enhance our understanding of name collisions that resulted from adding new gTLDs.

Would you be willing to please fill out the following brief survey regarding your experience? We would be grateful for your input!

https://forms.gle/1kj6VtEK1M5ANq8JA

Your responses will remain anonymous, and all personal information will be discarded after the research has concluded.

If you have any questions or would like to opt out of future communications related to this topic, please reply to this email.

Thank you for your help!

Sincerely,

Casey Deccio
ICANN's Name Collisions Analysis Project