Name Collision Analysis Project Study Two Report

5 April 2024

1 Introduction

The Name Collision Analysis Project (NCAP) Study Two final report brings together the research and analysis of several past studies, three studies conducted by the NCAP Discussion Group (DG), and years of NCAP DG presentations and meetings that touch on the critical issues surrounding name collisions. This report takes the reader through the methodology and findings of the three research studies and the analysis of the DG's work activities. The conclusions from those studies provide guidance for the topics regarding name collisions that the ICANN Board laid out in the ICANN Board resolutions 2017.11.02.29-2017.11.02.31.[1]

The Domain Name System (DNS) has evolved since the last round of new gTLD delegations began in 2012. Changes include the use of new DNS transports (such as DNS-over-TLS, DNS-over-HTTPS, and DNS-over-QUIC), additional DNS privacy extensions (such as QNAME minimization and Oblivious DNS), and features that address both privacy and query volume, such as aggressive NSEC and local root instances. Additionally, the rise of global public DNS resolver services has resulted in the increased consolidation of query traffic seen at authoritative servers, including the root servers. The introduction and growing use of all of these technologies challenge the effectiveness of the methods and data sets traditionally used for name collision analysis. This has resulted in the need for new methods to help understand when and where name collisions occur.

This changing landscape, in combination with the research done since 2012 (see Section 1.2) and community feedback, resulted in the Board's resolutions requesting that the ICANN Security and Stability Advisory Committee (SSAC) provide more definitive guidance as to what should be the next steps for the applications requesting delegation of .corp, .home, and .mail, three of the top Collision Strings identified in the 2012 round of gTLD delegations. In addition to this specific guidance, the effort was also expected to address the prevention or mitigation of name collisions more broadly.

Since 2014, Controlled Interruption has been ICANN's sole mechanism to alert users and system administrators to potential name collision issues. Several reports, including the "Mitigating the Risk of DNS Namespace Collisions Final Report," a commissioned document by JAS Global Advisors (the "JAS Report") and the Root Cause Analysis as commissioned through NCAP Study Two, have found Controlled Interruption to be effective, as a preemptive alert to the issues posed by that delegation, in disrupting systems that might be impacted by the general availability of a new gTLD. However, this disruption has had an impact ranging from mild to severe on affected systems. These side effects have caused investigators to reevaluate the use of Controlled Interruption and to explore additional techniques for identifying and mitigating the risks of name collision. Furthermore, the DG also evaluated gaps in the availability and completeness of data

[1] See Approved Board Resolutions | Regular Meeting of the ICANN Board | 2 November 2017, https://www.icann.org/en/board-activities-and-meetings/materials/approved-resolutions-regular-meeting-of-the-icann-board-02-11-2017-en#2.a.rationale

used to identify name collisions. The result of these evaluations is a workflow that offers guidance to ICANN org and gTLD applicants on identifying name collisions and identification of some of the risks of name collision before granting the delegation of a proposed gTLD to a Registry Operator. Implementing the recommendations in this workflow as part of the new gTLD application process will provide some mitigation against consequences experienced by affected systems.

The proposed workflow and name collision analysis process for applied-for strings include several techniques for gathering relevant data (See Section 3.5). These methods vary both as far as what information they provide and what risks or challenges go along with using them. This continues the understanding from past analysis that the prevention or mitigation of name collisions is fundamentally an issue of risk management. This risk management approach is also critical to understanding the Findings and Recommendations in Sections 4 and 5, respectively.

This report cannot assess all risk factors, as some of the relevant risks are not technical or operational, which means it cannot provide final answers on what techniques should be applied or what the final outcome of analysis should be. There is an element of judgment in applying all of the Findings and Recommendations in Sections 4 and 5, respectively. The NCAP DG has provided facts and analysis within its remit and the understanding available to the participants. However, the purpose of this report is to provide advice that will be further refined by input from—and ultimately implemented by—other parties. The proposed Technical Review Team (TRT), as described later in this report, will be expected to provide some of that judgment. In some cases, where there might be unusual risks and limited opportunities for mitigation, that judgment may belong to the ICANN org and ICANN Board. In such cases, the Findings and Recommendations compiled by the NCAP DG will be useful as input to those decisions.

The first section (Section 1) of this report describes the background of the NCAP and the mandate set forth by the ICANN Board in 2017. It goes on to describe the background that informed the direction of Study Two; the methodology of the study group as a whole, including the timeline of research, community outreach, study group consensus; and the terminology necessary to have a common understanding of how these terms are used in this report.

Section 2 of this report summarizes the three studies included in Study Two. While additional research may provide more clarity on the root causes (identification of the risk) and challenges of identifying name collisions, the results of these studies provide information not previously understood and inform the findings and recommendations in Sections 4 and 5.

Section 3 captures the years of discussion held by the DG. The expertise within that group provided necessary background and lived experiences that informed the Findings in Section 4 and the Recommendations in Section 5.

Appendix 1 offers a revised definition of *name collision* and a revised scope of work for the NCAP Study Two DG. Appendix 2 takes the research described in Section 2 and offers detailed explanations and guidance related to notification and data generation methods.

Appendix 3 includes a proposed workflow that focuses on risk management and presents a sample Technical Review Team report that could be used as a starting point for what the TRT will do as it conducts the analysis of the Collision Assessments proposed by the workflow. Specifically, the sample report addresses the Board resolution that specifically asks for guidance with respect to evaluating the status of .corp, .home, and .mail.

Appendix 4 includes the two Root Cause Analysis reports in full. Finally, an analysis of public comments on the draft Study Two Report submitted for public comment from January 19 to February 28, 2024.

While this report is primarily intended as input to the ICANN Board, all parties interested in the future expansion of the gTLD space, from applicants to community groups, will find the material relevant to their efforts.

1.1 Scope of Study Two

The SSAC was tasked by the ICANN Board in resolutions 2017.11.02.29-2017.11.02.31 to address a set of questions related to name collision.[2] To fulfill the Board's request, the SSAC chartered the Name Collision Analysis Project and developed three studies to answer the Board's questions. Study One was authorized by the ICANN Board in March 2019 and was completed in July 2020.

On 17 June 2020, the final draft of the Study One report was published for public comment.[3] The report on this public comment recommended that Studies Two and Three should "not be performed as currently designed." The DG agreed with this assessment and revised the design of NCAP Study 2 to take into account the issues raised by NCAP Study 1. In February 2021, the Board directed the NCAP DG to proceed with Study Two as redesigned.[4]

The results of these modifications dramatically reduced the scope, level of effort, total costs, and resources to execute Study Two. The revised Study Two proposal therefore was limited to the following goals:

[2] See Approved Board Resolutions | Regular Meeting of the ICANN Board | 2 November 2017, https://www.icann.org/en/board-activities-and-meetings/materials/approved-resolutions-regular-meeting-of-the-ican-board-02-11-2017-en#2.a.rationale

[3] See Managing the Risks of Top-Level Domain Name Collisions: Findings for the Name Collision Analysis Project (NCAP) Study 1, https://www.icann.org/en/system/files/files/ncap-study-1-report-19jun20-en.pdf

[4] See Approved Board Resolutions | Regular Meeting of the ICANN Board | 25 March 2021, https://www.icann.org/en/board-activities-and-meetings/materials/approved-resolutions-regular-meeting-of-the-ican-board-25-03-2021-en#2.b

1. Understand the root cause of most name collisions
2. Understand the impact of name collisions

And the final tasks included:

Task	Steps	Responsible Party
Study of ICANN Collision Reports	Perform an analysis of ICANN Collision Reports to determine the underlying cause of these collisions.	Technical Investigator
	Produce a report on the results of the analysis.	
Impact and Data Sensitivity Analyses	Research the impact of collisions with regards to Root servers and Resolvers for .corp, .home and .mail.	DG and Technical Investigator (guided by the DG / Admin team)
	Research the impact of collisions with regards to Root servers and Resolvers for other selected strings.	
	Based on the above research, evaluate the effectiveness of using multiple sources of collision data with regards to assessing the impact of collisions.	
	Undertake a public consultation on the findings relative to .corp, .home and .mail.	
	Produce a report on the results of this work.	
Response to Board Questions Relating to Study Two	Respond to Board questions based on the results of the Study of ICANN Collision Reports and Impact and Data Sensitivity Analyses.	Discussion Group
	Produce a report on the responses to Board questions.	
Final Report	Produce the final report for Study Two	
	Undertake a public consultation on the draft version of this report	

Table 1: Tasks Issued to the NCAP DG following the Study Two Proposal

It was noted by the DG that an item was erroneously included in the "In scope but not intended to be the subject of data studies" Name Collision definition used for Study One and Study Two and was appropriately corrected (see Appendix 1).

1.2 Background and Related Work

With over a decade's worth of discussion regarding the issue of DNS name collision, there is a wealth of background material to draw from on the topic. The diagram below (Figure 1) shows a timeline view of all the events and publications described in this background section.

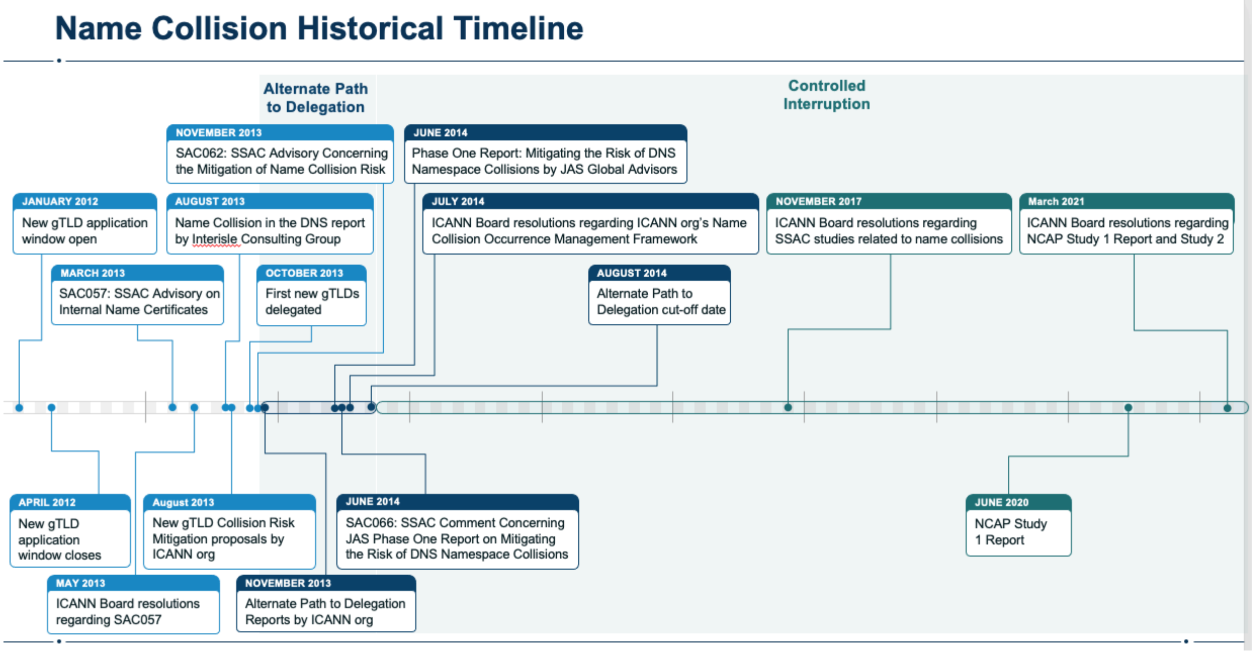

Name Collision Historical Timeline
Alternate Path to Delegation
Controlled Interruption
NOVEMBER 2013
SAC062: SSAC Advisory Concerning the Mitigation of Name Collision Risk
JUNE 2014
Phase One Report: Mitigating the Risk of DNS Namespace Collisions by JAS Global Advisors
JANUARY 2012
New gTLD application window open
AUGUST 2013
Name Collision in the DNS report by Interisle Consulting Group
JULY 2014
ICANN Board resolutions regarding ICANN org's Name Collision Occurrence Management Framework
NOVEMBER 2017
ICANN Board resolutions regarding SSAC studies related to name collisions
March 2021
ICANN Board resolutions regarding NCAP Study 1 Report and Study 2
MARCH 2013
SAC057: SSAC Advisory on Internal Name Certificates
OCTOBER 2013
First new gTLDs delegated
AUGUST 2014
Alternate Path to Delegation cut-off date
APRIL 2012
New gTLD application window closes
August 2013
New gTLD Collision Risk Mitigation proposals by ICANN org
JUNE 2014
SAC066: SSAC Comment Concerning JAS Phase One Report on Mitigating the Risk of DNS Namespace Collisions
JUNE 2020
NCAP Study 1 Report
MAY 2013
ICANN Board resolutions regarding SAC057
NOVEMBER 2013
Alternate Path to Delegation Reports by ICANN org

Figure 1: Name Collision Historical Timeline

Much of that material is captured in the NCAP Study One report, the ICANN Community Wiki, and the ICANN website. NCAP Study One provides an extensive, annotated bibliography of prior work related to name collisions, which we refer to in more detail below. The ICANN Wiki has a community-sourced page dedicated to name collisions that includes some history and enumeration of various events, as well as some references to notable material.[5] ICANN maintains a resource on its website called "Name Collision Resources & Information" with a broad set of materials applicable to the ICANN community, including a definition of name collisions.[6]

> A name collision occurs when an attempt to resolve a name used in a private name space[7] (e.g. under a non-delegated Top-Level Domain, or a short, unqualified name) results in a

[5] See ICANN Wiki: Name Collision, https://icannwiki.org/Name_Collision
[6] See ICANN, Name Collision Resources & Information, https://www.icann.org/resources/pages/name-collision-2013-12-06-en
[7] The reference text from which this quote was drawn writes the term "name space" as such.

> query to the public Domain Name System (DNS). When the administrative boundaries of private and public namespaces overlap, name resolution may yield unintended or harmful results.

We highlight some of the materials from these sources that significantly influenced this report.

1.2.1 SAC 057: SSAC Advisory on Internal Name Certificates

As the launch of the New gTLD Program was beginning, SSAC became aware of an issue with how *internal names* (which today we would compare to private use TLD strings) were being used in certificates and issued SAC057: SSAC Advisory on Internal Name Certificates.[8] This report included the first use of the term *name collision*, though it was not formally defined in that document.

On 18 May 2013, the ICANN Board adopted Resolutions 2013.05.18.08-2013.05.18.11 in response to SAC057, commissioning a study on the use of undelegated TLDs in enterprises.[9] This initial investigation into the risks and harms of name collisions occurred after the application period ended in April 2012. From there, the ICANN community continued to evolve the work as their understanding of the depth and breadth of the issue grew; ICANN org, in turn, continuously evolved the application evaluation workflow to account for the potential of name collisions.[10]

1.2.2 Name Collision in the DNS (the "Interisle Report")

The first publication within the ICANN context to directly address name collisions was an ICANN-commissioned report by Interisle Consulting Group, LLC, published on 2 August 2013.[11] Entitled "Name Collision in the DNS," (hereinafter referred to as the "Interisle Report") this was a study of the likelihood and potential consequences of a collision between new public gTLD labels and existing private uses of the same strings. This report established the first documented definition of a name collision:

> Name collision: two names that are represented by syntactically identical strings but belong to different semantic domains are said to "collide" when one of them appears in the other's semantic domain and is (mis)interpreted as if it belonged there.

[8] See SAC057: SSAC Advisory on Internal Name Certificates

[9] See Minutes | Regular Meeting of the ICANN Board | 18 May 2013, https://www.icann.org/en/board-activities-and-meetings/materials/minutes-regular-meeting-of-the-icann-board-of-directors-18-05-2013-en#2.a.rationale

[10] See ICANN Community Wiki: History of the Name Collision Analysis Project, https://community.icann.org/display/NCAP/History+of+the+Name++Collision+Analysis+Project.

[11] See Name Collision in the DNS, https://www.icann.org/en/system/files/files/name-collision-02aug13-en.pdf

The Interisle Report is used in this report as the baseline for comparison to all other work. The findings of the Interisle Report were primarily defined by the information that can be derived either directly or through analysis from the DNS request stream at the root servers that participated in the “Day in the Life of the Internet” (DITL) exercises organized by the DNS Operations, Analysis, and Research Center (DNS-OARC) in 2012 and 2013.

Among its many important insights are the following.

- The potential for name collisions is substantial and often arises from well-established policies and practices in private network environments.
- The delegation of almost any new TLD label would carry some risk of collision. The risk arises from the potentially harmful consequences of name collision, not the name collision itself.
- The designation of any applied-for string as “high risk” or “low risk” with respect to delegation as a new gTLD depends on both policy and analysis.
- The absence of evidence is not evidence of absence, i.e., even proposed new gTLD strings that appear to be “low risk” may be in widespread use on private networks.

1.2.3 New gTLD Collision Risk Mitigation

Building on this study, ICANN published its “New gTLD Collision Risk Mitigation” on 5 August 2013.[12] It included proposals to mitigate the collision risks between new gTLDs and existing private uses of the same strings. The proposals require the strings to be categorized according to their risk profile using the methodology described in the Interisle Report. The three proposals can be characterized as follows.

- For strings with a low-risk profile, the registry operator would deploy an authoritative name server for the TLD with an empty zone. For a period of not less than 30 days, the registry operator would be required to investigate all DNS queries received, contacting the source of the query and notifying that source of the imminent name collision that may result. The report noted the existence of recursive resolvers that would prevent the registry operator from seeing the actual source of the query; the mitigation proposal, therefore, included the requirement that registry operators obtain the cooperation of those recursive resolvers to identify the actual source of the query.
- For strings with a high-risk profile, the registry operator would need to demonstrate that the name collision could be mitigated such that the risk profile could be reduced to a low-risk profile. The low-risk profile mitigation proposal would then apply.

[12] See New gTLD Collision Risk Mitigation, https://www.icann.org/en/system/files/files/new-gtld-collision-mitigation-05aug13-en.pdf

- For strings with an uncalculated-risk profile, ICANN would conduct an additional study to assess the risk and understand what mitigation measures may be needed to allow these strings to move forward.

1.2.4 SAC062: SSAC Advisory Concerning the Mitigation of Name Collision Risk

On 7 November 2013, SSAC published SAC062, "SSAC Advisory Concerning the Mitigation of Name Collision Risk," establishing its first definition of a name collision.

> In the context of top level domains, the term "name collision" refers to the situation in which a name that is properly defined in the global Domain Name System (DNS) namespace (defined in the root zone as published by the root management partners - ICANN, U.S. Dept. of Commerce National Telecommunication Information Administration (NTIA), and VeriSign) may appear in a privately defined namespace (in which it is also syntactically valid), where users, software, or other functions in that domain may misinterpret it.[13]

SAC062 presented advice based on SSAC's review of the issues identified in the Interisle Report and ICANN's proposals to mitigate potential collision risks. SSAC's recommendation at the time was that high-risk strings should be considered for permanent reservation for internal or private use, suggesting that high-risk should include strings with documented evidence of broad and significant private usage. That definition could reasonably be expected to include .home and .corp, and perhaps .mail, since the volume of DNS query data did suggest significant private usage.

The SAC062 report defines an action called "trial delegation," which is similar to the Controlled Interruption that was ultimately deployed with a few critical differences.

- SAC062 defines two types of trial delegation: "DNS Infrastructure Testing" and "Application and Service Testing and Notification".
    - "DNS Infrastructure Testing" was characterized by the delegation of the prospective TLD string with an empty zone for the purpose of collecting data on the DNS queries received at the authoritative server for the TLD.
    - "Application and Service Testing and Notification" was characterized by the delegation of the prospective string with a wildcard resource and having it respond with synthesized responses for the purpose of causing a name collision

[13] See SAC062: SSAC Advisory Concerning the Mitigation of Name Collision Risk, https://www.icann.org/en/system/files/files/sac-062-en.pdf

and providing an opportunity to alert the client of the issue in a manner appropriate for the protocol (i.e., not just the DNS protocol) in use.

- The report further notes that if ICANN operated the trial delegation, “it would presumably be easier to quickly reverse the delegation if a significant consequence is discovered that required immediate mitigation.”

1.2.5 New gTLD Collision Occurrence Management Proposal

SAC062 was followed by ICANN’s publication of “New gTLD Collision Occurrence Management Proposal” to manage the collision occurrences between new gTLDs and existing private uses of the same strings.”[14] The Board approved this proposal for implementation and outreach via resolutions 2013.10.07.NG01 - 2013.10.07.NG02.[15] It includes the following definition of a name collision:

> “A name collision occurs when users unknowingly access a name that has been delegated in the public DNS when the user’s intent was to access a resource identified by the same name in a private network.”

Among the actions presented are the following.

- The Board deferred the delegation of .home, .corp, and .mail indefinitely and directed ICANN org to collaborate with the technical and security community to continue to study the issues presented by these strings.

- The Board further directed ICANN org to commission a study to develop a name collision occurrence management framework. The framework would specify a set of name collision occurrence assessments and corresponding mitigation measures[16], if any, that ICANN or TLD applicants may need to implement per second level domain name (SLD) seen in the DITL and other relevant datasets. The proposed name collision management framework will be made available for public comment.

- The proposal defined a “Collision Occurrence Assessment” that ICANN would conduct and deliver to each applicant and make available to the community. This assessment would include suggested mitigation methods, among which was the option to implement

[14]See New gTLD Collision Occurrence Management, https://www.icann.org/en/system/files/files/resolutions-new-gtld-annex-1-07oct13-en.pdf

[15] See Approved Resolutions | Meeting of the New gTLD Program Committee | 7 October 2013, https://www.icann.org/en/board-activities-and-meetings/materials/approved-resolutions-meeting-of-the-new-gtld-program-committee-07-10-2013-en#1.a

[16] From New gTLD Collision Occurrence Management Proposal: “Note that measures taken by ICANN or TLD applicants are attempts to mitigate unintended consequences or harm by preventing a name collision from occurring. These measures do not mitigate the causes of collision occurrences. Mitigating causes is a matter for users, private network operators, software developers, or equipment manufacturers to address.”

a trial delegation of some form. Details of the proposed methods can be found in Section 3.2 of the proposal.

- Section 3.3 of the proposal defined a mitigation measure called “Alternate Path to Delegation.” This required registry operators to “block” the use of an extensive set of potential second-level domain names (SLDs). This was done to ensure that a client attempting to use the domain name that would result in a name collision would continue to receive a DNS response indicating the name did not exist. Understanding that requirement is critical to the NCAP Study Two report.
- Section 3.4 empowered ICANN to develop an outreach campaign to raise general awareness and provide advice to minimize the potential for unintended consequences or harm.

ICANN completed the “Collision Occurrence Assessment”, using DITL and other relevant data as an input, for all applied-for strings on 17 November 2013 and published them as “Reports for Alternate Path to Delegation Published”.[17] This assessment found 25 strings ineligible for the Alternate Path to Delegation, .mail among them. These strings would have to wait for the name collision management framework to be developed. The strings .home and .corp, which the Board had indefinitely deferred, were also excluded. All others could proceed to implement the Alternate Path to Delegation if they were approved for delegation and the corresponding registry operator chose to do so. According to ICANN’s Delegated Strings page, 370 TLDs were delegated via the Alternate Path to Delegation.[18]

1.2.6 Name Collision Occurrence Management Framework

On 4 June 2014, ICANN published the Phase One Report, "Mitigating the Risk of DNS Namespace Collisions,"[19] a commissioned report by JAS Global Advisors (hereinafter described as the “JAS Report”); the final report was published in 2015.[20] ICANN used the JAS Report, which primarily relied upon DITL data analysis, to develop the “Name Collision Occurrence Management Framework[21],” a guide for ICANN and the new gTLD registry operators on how to handle name collisions. The report includes several recommendations immediately relevant to

[17] See ICANN New Generic Top-Level Domains: Reports for Alternate Path to Delegation Published, https://newgtlds.icann.org/en/announcements-and-media/announcement-2-17nov13-en

[18] See ICANN New Generic Top-Level Domains: Delegated Strings, https://newgtlds.icann.org/en/program-status/delegated-strings

[19] See Mitigating the Risk of DNS Namespace Collisions: Phase One Report, https://www.icann.org/en/system/files/files/name-collision-mitigation-26feb14-en.pdf

[20] See Mitigating the Risk of DNS Namespace Collisions: Final Report, https://www.icann.org/en/system/files/files/name-collision-mitigation-final-28oct15-en.pdf

[21] See ICANN Name Collision Occurrence Management Framework, https://www.icann.org/en/system/files/files/name-collision-framework-30jul14-en.pdf

the Study Two report; we refer the reader to the JAS Report for the supporting analysis associated with each recommendation.

- **Recommendation 1**: The TLDs .corp, .home, and .mail be referred to the Internet Engineering Task Force (IETF) for potential RFC 1918-like protection/treatment.
- **Recommendation 3**: Emergency response options are limited to situations where there is a reasonable belief that the DNS namespace collision presents a clear and present danger to human life.
- **Recommendation 4**: Root-level de-delegation of a production TLD is not considered as an emergency response mechanism under any circumstances.
- **Recommendation 5**: ICANN leverage the EBERO mechanisms and functionality to respond to DNS namespace-related issues.
- **Recommendation 6**: ICANN require new TLD registries to publish the controlled interruption zone immediately upon delegation in the root zone. After the 90-day period, there shall be no further collision-related restrictions on the registry.
- **Recommendation 10**: ICANN work with the IETF to identify a mechanism for IPv6 that provides similar functionality to that available in IPv4's "localhost" reserved prefix.
- **Recommendation 14**: ICANN request that the appropriate bodies further explore issues relating to collisions in existing DNS namespace, the practice of "domain drop catching," and the associated data feeds that may be leveraged by attackers when attempting to exploit collisions.

1.2.7 SAC066: SSAC Comment Concerning JAS Phase One Report on Mitigating the Risk of DNS Namespace Collisions

On 6 June 2014, SSAC published SAC066, "SSAC Comment Concerning JAS Phase One Report on Mitigating the Risk of DNS Namespace Collisions."[22] In that document, SSAC reviewed the Phase One Report by JAS Global Advisors noted in the previous paragraph. SAC066 used the following definition of a name collision in its report:

> The term 'name collision' refers to the situation where a name that is defined and used in one namespace may also appear in another. Users and applications intending to use a name in one namespace may actually use it in a different one, and unexpected behavior may result where the intended use of the name is not the same in both namespaces.

[22] See SAC066: SSAC Comment Concerning JAS Phase One Report on Mitigating the Risk of DNS Namespace Collisions, https://www.icann.org/en/system/files/files/sac-066-en.pdf

SSAC identified eight issues with the Phase One JAS Report and made a recommendation about each of them. These include:

- ICANN should perform an evaluation of potential notification approaches against at least the requirements provided by the SSAC prior to implementing any notification approach.
- ICANN should implement a notification approach that accommodates Internet Protocol Version 6 (IPv6)-only hosts as well as IP Version 4 (IPv4)-only or dual-stack hosts.
- ICANN should seek to provide stronger justification for extrapolating findings based on one kind of measurement or data gathering to other situations.

1.2.8 Name Collision Occurrence Management Framework

Finally, we have the current "Name Collision Occurrence Management Framework,"[23] originally published on 30 July 2014 and approved and directed for implementation by the ICANN Board with Resolution 2014.07.30.NG01[24]. This framework has remained in force since it was published and is the current mechanism through which ICANN assesses name collisions. ICANN considered the recommendations in the JAS Report and the advice in SAC062 and SAC066. The Framework begins with the following definition of a name collision:

> A name collision occurs when a user unknowingly accesses a name that has been delegated in the public DNS when the user's intent is to access a resource identified by the same name in a private network. Circumstances like these, where the administrative boundaries of private and public namespaces overlap and name resolution yields unintended results, present concerns and should be avoided if possible.

Key elements of the Name Collision Occurrence Management Framework's methodology include:

- Registry operators are required to act on name collision reports forwarded by ICANN within two hours of receipt.
- Controlled Interruption, as described by the JAS Report, is required of all new gTLDs, notably because it was decided its good notification features combined with its superior privacy protection were preferred to the use of a honeypot as defined by the SSAC.
- The lack of IPv6 support was accepted as a tolerable risk; while recognized as a gap, it was not described as a blocking concern. The Framework instead suggested that ICANN

[23] See Name Collision Occurrence Management Framework, https://www.icann.org/en/system/files/files/name-collision-framework-30jul14-en.pdf

[24] See Approved Resolutions | Meeting of the New gTLD Program Committee | 30 July 2014, https://www.icann.org/en/board-activities-and-meetings/materials/approved-resolutions-meeting-of-the-new-gtld-program-committee-30-07-2014-en#1.a

"will work within the IETF and with other relevant technical communities to identify a mechanism for IPv6 that provides similar functionality to that available in IPv4's "Loopback" reserved prefix.

- Registry operators agree that ICANN may designate an Emergency Back-End Registry Operator (EBERO) if the Registry Operator is unable or unwilling to comply with a measure to avoid harm from name collision in a timely manner.
- The recommendation in the JAS Report to treat .mail the same as .home and .corp was accepted by ICANN, i.e., the delegation of .mail was deferred indefinitely.
- ICANN will produce information materials as needed regarding name collision.
- ICANN will limit emergency response for name collision reports to situations where there is a reasonable belief that the name collision presents a clear and present danger to human life.

1.2.9 SSAC Proposals for the Name Collision Analysis Project

Moving ahead to 2017, the ICANN Board requested that SSAC conduct studies to present a data analysis on available information and provide advice to the Board on the topics around DNS name collision.[25] The details of the resolutions and the embedded questions are covered later in this report. Two key elements from those resolutions are that SSAC was asked to propose a proper definition of a name collision and that the Board defined a new term, Collision String, as a category for undelegated strings that should be considered strings that manifest name collisions.

In response, the SSAC proposed the "Name Collision Analysis Project (NCAP)," which was quite broad and consistent with SSAC's prior advice on the issue of name collisions.[26] The final SSAC NCAP Proposal, published in September 2018, was organized into three studies.[27] In broad terms, the purposes were:

> **Study One**: To establish a shared understanding of what we know about name collisions and a data repository for studying them.
>
> **Study Two**: To conduct an analysis with the goals of understanding the source of name collisions and developing a sustainable framework for evaluating the risk of the manifestation of a name collision.

[25] See Approved Board Resolutions | Regular Meeting of the ICANN Board | 2 November 2017, https://www.icann.org/en/board-activities-and-meetings/materials/approved-resolutions-regular-meeting-of-the-icann-board-02-11-2017-en#2.a

[26] See ICANN Community Wiki: SSAC Name Collision Analysis Project (NCAP) Home, https://community.icann.org/display/NCAP/SSAC+Name+Collision+Analysis+Project+%28NCAP%29+Home

[27] See SSAC Proposal for the Name Collision Analysis Project

> **Study Three**: To study and propose mitigation and remediation strategies for responding to name collisions.

The ICANN Board accepted SSAC's suggestion for professional project management, and ultimately the project was assigned to ICANN's Office of the Chief Technology Officer (OCTO) to manage. OCTO reviewed SSAC's project proposal and, in collaboration with the SSAC, made minor revisions to the project and developed a budget. The ICANN Board approved moving forward with the Revised Study One[28] on 14 March 2019 with Resolutions 2019.03.14.20 – 2019.03.14.23.[29]

The revised proposal reduced the scope of Study One by removing the creation of the data repository and deferring that work until Study Two, thus reducing the duration and cost of the study. The proposal noted the following definition of a name collision as baseline input for the NCAP Project.

> Name Collision refers to the situation where a name that is defined and used in one namespace may also appear in another. Users and applications intending to use a name in one namespace may actually use it in a different one, and unexpected behavior may result where the intended use of the name is not the same in both namespaces. The circumstances that lead to a name collision could be accidental or malicious. In the context of top-level domains (TLDs), the conflicting namespaces are the global Internet Domain Name System (DNS) namespace reflected in the root zone as published by the Root Zone Management Partners and any other namespace, regardless of whether that other namespace is intended for use with the DNS or any other protocol.

The formation of the DG was announced on 17 April 2019, inviting anyone in the ICANN Community to join the DG.[30] The initial tasks of the DG were to define the term 'name collisions' to scope the material to be researched and review the Request For Proposal developed by OCTO seeking a contractor to complete the work. Ultimately, the goals of Study One were three-fold.

1. To produce a summary report on the topic of name collision that brings forth important knowledge from prior work in the area.

[28] See SSAC Proposal for the Name Collision Analysis Project (Revised by ICANN Office of the CTO)

[29] See Minutes | Regular Meeting of the ICANN Board | 14 March 2019, https://www.icann.org/en/board-activities-and-meetings/materials/minutes-regular-meeting-of-the-icann-board-14-03-2019-en#2.h.1

[30] See Project Overview for the Name Collision Analysis Project (NCAP) Study 1: Request for Proposal, https://www.icann.org/en/system/files/files/rfp-ncap-study-1-09jul19-en.pdf

2. To create a list of datasets used in past name collision studies; identify gaps[31], if any; and make a list of any additional datasets required to complete Studies Two and Three successfully.

3. To offer a recommendation on whether Studies Two and Three should be performed based on the results of the survey of prior work and the availability of datasets.

The final Study One Report[32] was published on 19 June 2020 and included four (4) significant findings, excerpted here from the Executive Summary.

1. Name collisions have been a known problem for decades, possibly as early as the late 1980s. Reports, papers, and other work regarding name collisions were sparse and sporadic until 2012, at which point many organizations and individuals began publishing extensively on the topic. Workshops were held in 2013 and 2014. Since ICANN approved the Name Collision Occurrence Management Framework in 2014, which instituted controlled interruption as the mitigation strategy for new TLDs, the volume of work on name collisions by academic institutions, the security industry, IT product and service vendors, and others has greatly decreased. The only known work on name collisions during the past few years has been from ICANN by the NCAP DG and the New gTLD Subsequent Procedures (SubPro) Working Group. Since mid-2017, there has not been any published research into the causes of name collisions or new name collision mitigation strategies.

2. Since controlled interruption was instituted, there have been few instances of name collision problems being reported to ICANN or reported publicly through technical support forums, mailing lists, and other means. Most problems occurred during 2014, 2015, or 2016, with only a single problem reported to ICANN during the three-year period from 2017 through 2019, as well as a sharp dropoff in public reports during the same period. Only one of the reports to ICANN necessitated action by a registry, and none of the public reports surveyed mentioned major harm to individuals or organizations.

3. Prior work and name collision reports have indicated there are several types of root causes of name collisions – perhaps a dozen or more. These root causes have typically been found by individuals researching a particular leaked TLD to find its origin, not by examining datasets. There is unlikely to be any dataset that would contain root causes; identifying root causes is generally going to require research of each TLD involved in name collisions on a case-by-case basis.

[31] From Project Overview for the Name Collision Analysis Project (NCAP) Study 1: Request for Proposal: “Gaps in the data refers to types, sources, specific events captured, etc., that were not used in prior work but would have been useful or even necessary for the prior work to have been comprehensive.”

[32] See Managing the Risks of Top-Level Domain Name Collisions: Findings for the Name Collision Analysis Project (NCAP) Study 1, https://www.icann.org/en/system/files/files/ncap-study-1-report-19jun20-en.pdf

4. No gaps or other issues have been identified in accessing the datasets that would be needed for Studies Two and Three.

The final report also made a significant recommendation regarding the execution of NCAP Studies Two and Three, that Studies Two and Three should *not* be performed as currently designed. The Study One Report Executive Summary continued as follows.

> Recent discussions among NCAP DG members indicate differences of opinion as to whether controlled interruption has been "successful." It does not appear that criteria for success are formally defined, and until such criteria are defined, disagreements are likely to continue. That being said, however, there have been minimal name collision problems reported since controlled interruption was instituted, given the number of new TLDs it has been used for in the past six years. Research conducted for this report included extensive searches for evidence, and NCAP DG members were repeatedly asked to provide information on any evidence they were aware of. The counterargument to this has been the old saying, "Absence of evidence is not evidence of absence." Although that saying has merit, over time the continued absence of evidence that controlled interruption has not been successful makes it less likely to be true. The lack of interest in alternatives to controlled interruption outside a few groups within ICANN further supports the likelihood that controlled interruption has been successful.
>
> Given these findings, the recommendation is that Studies 2 and 3 should not be performed as currently designed. Regarding Study Two, analyzing datasets is unlikely to identify significant root causes for name collisions that have not already been identified. New causes for name collisions are far more likely to be found by investigating TLD candidates for potential delegation on a case by case basis. Regarding Study 3, controlled interruption has already proven an effective mitigation strategy, and there does not appear to be a need to identify, analyze, and test alternatives for the vast majority of TLD candidates.
>
> All of that being said, this does not necessarily mean further study should not be conducted into name collision risks and the feasibility of potentially delegating additional domains that are likely to cause name collisions. Most notably, the Study 3 question of how to mitigate name collisions for potential delegation of the .corp, .home, and .mail TLDs is still unresolved. However, the proposals for Studies 2 and 3, which were developed years ago, do not seem to be effective ways of achieving the intended goals.

SSAC agreed with the assessment regarding Studies Two and Three as currently designed and set to work reframing Study Two and working with OCTO, as the Project Manager, to prepare a budget; Study Three would be reconsidered after Study Two completed[33]. On 5 February 2021,

[33] Upon completing Study Two, the NCAP DG recommends that ICANN not move ahead with Study Three (See Recommendation 11)

the SSAC submitted a Revised Proposal for Study Two[34] to the ICANN Board. On 25 March 2021, the ICANN Board accepted the Study One final report, approved the Revised Proposal for Study Two, and directed the DG to proceed with the Revised Study Two with Resolutions 2021.03.25.11 – 2021.03.25.14.[35] Readers are referred to the revised proposal for a discussion of the detailed changes from the original proposal. The revised Study Two, for which this report is the final work product, stated four (4) objectives:

- Perform a study of ICANN Collision Reports.
- Perform Impact and Data Sensitivity Analyses with respect to name collisions.
- Respond to Board Questions Relating to Study Two.
- Produce a final report on Study Two.

1.2.10 Final Report on the new gTLD Subsequent Procedures Policy Development Process

Overlapping the efforts of Study One and Study Two is the output of the ICANN Subsequent Procedures (SubPro) Working Group, which published its final report on 1 February 2021, Final Report on the new gTLD Subsequent Procedures Policy Development Process.[36] In Topic 29 of that report, the working group focused entirely on the issue of name collisions. They offered a recommendation, several affirmations, and implementation guidance to ICANN org on how to identify and mitigate name collisions before the next round of gTLDs. Readers of this report are encouraged to review the detailed rationale and support for the recommendation, affirmations, and implementation guidance in the final SubPro report. As these are both relevant and important to the NCAP work, their summary is excerpted here for easy reference.

> **Recommendation 29.1**: ICANN must have ready prior to the opening of the Application Submission Period a mechanism to evaluate the risk of name collisions in the New gTLD evaluation process as well as during the transition to delegation phase.
>
> **Affirmation 29.2**: The Working Group affirms continued use of the New gTLD Collision Occurrence Management framework unless and until the ICANN Board adopts a new mitigation framework. This includes not changing the controlled interruption duration and the required readiness for human-life threatening conditions for currently delegated gTLDs and future new gTLDs.

[34] See SSAC 2021-02: Revised Study Two Proposal for the Name Collision Analysis Project, https://itp.cdn.icann.org/en/files/security-and-stability-advisory-committee-ssac-correspondence/ssac2021-02-05feb21-en.pdf

[35] See Approved Board Resolutions | Regular Meeting of the ICANN Board | 25 March 2021, https://www.icann.org/en/board-activities-and-meetings/materials/approved-resolutions-regular-meeting-of-the-icann-board-25-03-2021-en#2.b

[36] See Final Report on the new gTLD Subsequent Procedures Policy Development Process, https://gnso.icann.org/sites/default/files/file/field-file-attach/final-report-newgtld-subsequent-procedures-pdp-02feb21-en.pdf

> **Implementation Guidance 29.3**: To the extent possible, ICANN should seek to identify high-risk strings in advance of opening the Application Submission Period, which should constitute a "Do Not Apply" list. ICANN should also seek to identify aggravated risk strings in advance of the next application window opening and whether it would require a specific name collision mitigation framework.
>
> **Implementation Guidance 29.4**: To the extent possible, all applied-for strings should be subject to a DNS Stability evaluation to determine whether they represent a name collision risk.
>
> **Implementation Guidance 29.5**: The ICANN community should develop name collision risk criteria and a test to provide information to an applicant for any given string after the application window closes so that the applicant can determine if they should move forward with evaluation.
>
> **Implementation Guidance 29.6**: If controlled interruption (CI) for a specific label (usually a 2nd-level domain) is found to cause disruption, ICANN may decide to allow CI to be disabled for that label while the disruption is fixed, provided that the minimum CI period is still applied to that label.

1.3 Methodology

With the acceptance of the revised Study Two proposal, the DG commenced the proposed studies and began meeting regularly to discuss progress and direction. While the DG considered the questions assigned by the ICANN Board, the researchers collected and analyzed available data relevant to understanding how to observe and measure the impact of name collisions; each report describes its specific methodology.

The DG chairs called for consensus on the responses to the Board questions, the study reports, and any special terminology after the discussion on each item was concluded during the regular conference calls. Two of the study reports went out for public comment prior to their being used in this report to finalize the findings and recommendations to the ICANN Board. The NCAP project was also presented at ICANN74[37], ICANN75[38], ICANN76[39], ICANN77[40] and ICANN78[41] to ensure the broader community was aware of the work, findings, and pending recommendations.

[37] See ICANN74: NCAP Status Update, https://74.schedule.icann.org/meetings/wcin8eB2MQNNRwWP6

[38] See ICANN75: NCAP Final Update: Preparation for Public Comment, https://75.schedule.icann.org/meetings/WxsCLa9h4NapEaq6n

[39] See ICANN76: Name Collision Analysis Project (NCAP): Study 2 Update, https://icann76.sched.com/event/1IfwG/name-collision-analysis-project-ncap-study-2-update

[40] See ICANN77: Name Collision Analysis Project Study 2 Update, https://icann77.sched.com/event/1N5ZJ/name-collision-analysis-project-study-2-update

[41] See ICANN78: Name Collision Analysis Project Study 2 Update, https://icann78.sched.com/event/1Sgpj/name-collision-analysis-project-study-2-update

1.4 Terminology

- Allocation - The process by which the Board decides whether to allow an applied-for TLD to be granted to the applicant.
- Collision String - A string that manifests Name Collisions (see "Name Collision" below)
  - Collision String List - a list of names not to be allocated nor delegated (on the Collision String List).[42]
- Controlled Interruption - "Controlled interruption is a method of notifying system administrators who have configured their networks incorrectly (knowingly or unknowingly) of the namespace collision issue, and helping them mitigate potential issues."[43]
- Critical Diagnostic Measurement - properties that help determine the scope, impact, and potential harm of name collisions
- Day-In-The-Life (DITL) - a large-scale data collection project run by DNS-OARC[44] undertaken every year since 2006.
- Delegation - the technical process of creating a subdomain; in the context of ICANN's responsibility for DNS, it means creating a new subdomain to the DNS root zone[45]. Such a name is a "TLD"; it's a subdomain of the root, and in turn delegates second or lower level names to registrants. This should be explicitly distinct from the process of granting the TLD to an applicant. (See "Allocation" above.)
- Grant - the administrative process of approving an application for a new TLD to a registry operator
- Harm - may include numerous things, from cybersecurity risks to reputational damage to physical impacts, making it difficult to appropriately apply scale and context to this otherwise broad term within the scope of name collisions. The DG's definition of harm is provided in the subsection that follows (See Section 1.4.1).
- Name Collision - (used in Study One and RFP) Name collision "refers to the situation where a name that is defined and used in one namespace may also appear in another. Users and applications intending to use a name in one namespace may attempt to use it in a different one, and unexpected behavior may result where the intended use of the name is not the same in both namespaces. The circumstances that lead to a name collision could be accidental or malicious."
- Name Collision Occurrence Assessment - formal output of the Technical Review Team

[42] See Proposed Definition of Name Collisions and Scope of Inquiry for the Name Collisions Analysis Project, https://www.icann.org/en/public-comment/proceeding/proposed-definition-of-name-collisions-and-scope-of-inquiry-for-the-name-collisions-analysis-project-02-07-2019

[43] See ICANN Frequently Asked Questions: Name Collision Occurrence Management Framework for Registries, https://www.icann.org/resources/pages/name-collision-ro-faqs-2014-08-01-en

[44] See Domain Name System Operations Analysis and Research Center (DNS-OARC): DITL, https://www.dns-oarc.net/oarc/data/ditl

[45] See ICANN Principles for Delegation and Administration of ccTLDs Presented by Governmental Advisory Committee, http://archive.icann.org/en/committees/gac/gac-cctldprinciples-23feb00.htm

- Query Volume - The number of DNS requests received for a string.
- Root Server Identity (RSI) - 13 identities, each of which is named with the letters ‘a’ to ‘m’, collectively administered by twelve root server operators. They are named in the ‘root-servers.net’ domain. Each root server identity is implemented by multiple separate servers.
- Search List Processing - “A Domain Name System (DNS) “search list” (hereafter, simply “search list”) is conceptually implemented as an ordered list of domain names. When the user enters a name, the domain names in the search list are used as suffixes to the user-supplied name, one by one, until a domain name with the desired associated data is found or the search list is exhausted.” [46]
- Source Diversity - The number of distinct source IP addresses, distinct /24 or /48 IP blocks, and/or distinct number of ASNs requesting a string. This results in three different measurements/numbers used in DNS query analysis.
- Risk - The report doesn’t recommend a specific or formalized risk management approach and uses this term in its “plain language” meaning to refer to the possibility of adverse outcomes from an action or a decision. In this context, an essential component of the DG’s approach to name collision analysis and mitigation is that any decision or course of action can have negative outcomes, and much of the work of name collision analysis is in determining the likelihood of different impacts and tradeoffs between possible benefits and harms.

1.4.1 Impact and Harm

The JAS Report described several of the challenges of enumerating harm when it comes to name collisions. Arguments around concepts of national security, economic hardship, and adherence to the law are impossible to manage in a diverse global context. Their final recommendation on the topic was:

> As such, we recommend that emergency response be limited to scenarios where there is a reasonable belief that the DNS namespace collision presents a clear and present danger to human life.[47]

The NCAP DG felt it necessary to extend the discussion of harm to include its potential. As noted in response to the Board questions, the DG approached harm as follows:

> To address the Board’s question, the discussion group focused on three aspects of harm: potential harm, reported harm, and systemic harm. Potential harm is a set of circumstances that might lead users and systems to be negatively impacted by name

[46] See SAC064: SSAC Advisory on DNS “Search List” Processing, https://www.icann.org/en/system/files/files/sac-064-en.pdf

[47] See Mitigating the Risk of DNS Namespace Collisions: Final Report (“JAS Report”), https://www.icann.org/en/system/files/files/name-collision-mitigation-final-28oct15-en.pdf

> collisions, with their possible levels of impact. Reported harm is based on actual experience disclosed by organizations and individuals impacted by name collisions. Systemic harm is a broader concern which the Board must consider if the risk of name collisions damages the reputation and ability to trust the responses for names in the DNS.[48]

The Board should also consider the question of harm from a more systemic perspective. If harm from name collisions becomes a common occurrence, then trust in the DNS as a whole is lost. This is discussed further in the DGs consideration of harm in the response to the Board Questions. When considering the risk of name collisions, the potential for harm must be part of the risk assessment. Ultimately, the goal is to prevent reported harm by evaluating the potential and reacting accordingly.

[48] See Responses to Board Resolution 2017.11.02.30 for Name Collision Analysis Project Discussion Group: "Theme 3: Harm," published as part of the 19-January-2024 NCAP Study Two Report Public Comment Proceeding

2 Overview of NCAP Study Two Reports

As described in its revised scope, the NCAP DG conducted three studies as part of Study Two:

- Case Study of Collision Strings[49]
- A Perspective Study of DNS Queries for Non-Existent Top-Level Domains[50]
- Root Cause Analysis: wpad.domain.name[51] and New gTLD Collisions[52]

Each study offered several insights into how to look for and understand the impact of name collisions.

The first study report, the Case Study of Collision Strings, helped define all the Critical Diagnostic Measurements (CDMs) required to identify name collisions and, further, how to assess the impact of a name collision.

The second study report, A Perspective Study of DNS Queries for Non-Existent Top-Level Domains, considered if and how the available data sets from both individual root servers and global public resolvers were representative or not of the overall picture of the DNS queries that would help identify name collisions.

The root cause analysis resulted in two reports, both investigating submissions to ICANN related to name collisions experienced. The first, Root Cause Analysis - `wpad.domain.name`, investigates reports of exploits associated with the domain name wpad.domain.name in connection with home routers. The second, Root Cause Analysis - New gTLD Collisions, provides both a quantitative analysis, using historical DNS query data, as well as a qualitative analysis, using submitted name collision reports and results from a name collision survey. It includes assessments of the pervasiveness of private use of newly-delegated TLDs in DNS suffixes, the effectiveness of controlled interruption in notification and root cause identification, the severity of impact felt by affected parties, and anecdotal configurations that were common causes of name collisions.

The following sections describe the results of those studies in greater detail.

2.1 Case Study of Collision Strings

The DG met over the course of approximately two years to evaluate and consider topics posed by the ICANN Board on the delegation of indefinitely deferred TLDs .corp, .home, and .mail. The

[49] See Case Study of Collision Strings, https://www.icann.org/en/system/files/files/case-study-collision-strings-13jul22-en.pdf

[50] See A Perspective Study of DNS Queries for Non-Existent Top-Level Domains (Previously termed “Impact and Data Sensitivity Analysis”), https://www.icann.org/en/system/files/files/perspective-study-dns-queries-non-existent-top-level-domains-13jul22-en.pdf

[51] See Appendix 4a: Root Cause Analysis - wpad.domain.name

[52] See Appendix 4b: Root Cause Analysis - New gTLD Collisions

group undertook a review of past studies and literature and conducted its own analysis from two root server identities. The result of that review is a modern picture of the impact and potential harm due to name collisions with the undelegated names under study. The analysis provides a sufficient basis from which to draw a number of important findings. One such finding is the observation that queries for these undelegated names are increasing in both volume and diversity. These facts suggest that challenges relating to impact and risk are also increasing. The group also identified a number of Critical Diagnostic Measurements that help determine the scope, impact, and potential harm of name collisions.

2.2 A Perspective Study of DNS Queries for Non-Existent Top-Level Domains

The report's analysis shows that no view at a single root server is comprehensive. However, when considering DNS clients that meet a defined query rate, a single root server observes query traffic from about two-thirds of resolvers that are observed across the entire system. Additionally, there are notable differences in DNS traffic observed by recursive resolvers and at the root server system. These findings are significant in terms of how future guidance and advice may be applied to name collision risk assessments. Specifically, these perspective differences affect the effectiveness of top-N lists, particularly when they are generated from a single source.

The publication of top-N lists of non-existent TLDs can make applicants aware of strings that exhibit some risk associated with name collisions. However, the effectiveness of such lists is limited. The very fact that these lists contain only the top N, ranked by some criteria, is constraining. This is particularly so when they are generated only from a single data source (e.g., root server queries or a single recursive resolver or at a single point in time). Because there are multiple perspectives in the DNS ecosystem, the absence of a string on a top-N list does not provide any assurance the string is void or absent of name collision risks, nor does the magnitude or ranking of a string that does show up in the list. For example, this analysis shows that non-existent TLDs observed at high volumes by some recursive resolvers are not seen in the same rankings by root servers.

2.3 Root Cause Analysis Reports

The motivation for the root cause analysis was to investigate the name collision reports submitted to ICANN to better understand what caused the name collisions, their severity, and the effectiveness of controlled interruption. Beginning with those name collision reports, a systematic and comprehensive study of name collisions associated with the delegation of new TLDs since the introduction of controlled interruption was undertaken. The study incorporates five (5) data sets:

- the 47 name collision reports submitted via ICANN's name collisions Web submission form;
- historical DNS query data extracted from passive DNS observation from the time of delegation of each of the 885 TLDs delegated since August 2014.
- root DNS query data from the 48-hour once-yearly day-in-the-life (DITL) collection from 2014 to 2021;
- results from a Web search for "127.0.53.53"; and
- responses from a name collisions survey sent to both a general technical audience and those inferred to have been affected by name collisions.

Key findings from the research and analysis of available data include:

- The private use of DNS suffixes is widespread.
- The name collision reports are supported strongly by measured data.
- The usage of private DNS suffixes colliding with newly-delegated TLDs has decreased over time.
- Controlled interruption is effective at disruption but not at root cause identification.
- Configuring DNS resolvers as authoritative for DNS suffixes is not a panacea.
- The impact of TLD delegation ranged from no impact to severe impact.
- The respondents' response to controlled interruption was overall neutral.
- Name collisions were diverse, both in terms of the application involved and their root causes.

Seven of the reports submitted via ICANN's name collisions report form were related to the interception of user Web traffic due to the combination of systems that use the Web Proxy Auto-Discovery protocol (WPAD), inadvertent usage of the domain name 'domain.name' in home router software, and the delegation of wpad.domain.name in the public DNS. While these issues do not fit in the same category as name collisions at the TLD level, the largest constituency of reports submitted to ICANN were associated with this issue. Thus, the DG agreed to additional research in a root cause analysis specific to .WPAD. This research contains a full delegation and resolution history of wpad.domain.name, an analysis of related queries observed at the root servers, and a behavioral analysis of the services operated by wpad.domain.name, i.e., what privacy and operability concerns might have been encountered by affected users.

For more detail on these findings, please review the Root Cause Analysis reports in Appendix 4.

3 Summary of NCAP Discussion Group Activities

The study reports described above, combined with a review of the materials gathered in Study One and a review of the evolution of the DNS and Internet infrastructure since the last round of new gTLDs, provided a foundation for consideration of name collisions today as compared to the last round and the opportunity to reconsider how to examine the risk they present to the security and stability of the DNS. In addition, while the prior reports focused on available data, the discussions of the DG worked to put that information, and more, in context.

3.1 The NCAP Gap Analysis

NCAP Study One offered an in-depth review of prior work around identifying and handling name collisions. Between the publication of the NCAP Study One report[53] and the Board resolutions 2021.03.25.11 – 2021.03.25.14[54] that approved the revised proposal for Study Two, members of the group focused their efforts on identifying the gaps between the technology that uses the DNS and the mechanisms used to identify and assess name collision risks. That effort informed the Revised Study Two Proposal[55].

The NCAP Gap analysis offered both hypotheses to be tested and baseline assertions to inform the direction of work for Study Two and were included in Appendix 2 of the Revised Study Two Proposal. The substantive text is included here for ease of reference.

> 1) <u>Data Sets:</u> Since the new gTLD program, various new data sets have become available that may provide additional telemetry to better understand and assess name collision risks. The new gTLD name collision risk assessment was conducted against a few years of Day In the Life of the Internet (DITL) DNS traffic data. Unfortunately, the DITL data set has several limitations, as it only provides a few days per year of authoritative root server DNS traffic, is contributed by root server operators on a voluntary basis, and may be anonymized due to privacy concerns. Since the last TLD round, the collection of DITL data has continued and may provide better longitudinal measurements pre/post the new TLD delegations. Other entities have also started to retain high fidelity root DNS traffic that may provide better insights. The emergence of popular open recursive resolvers has also transpired and dramatically shaped the DNS ecosystem since the new gTLD delegations. These recursive services may provide a richer and more complete understanding of name collisions if they can be utilized for analysis. Other potential data repositories of interest would also include the ORDINAL DNS data as well as Certificate Transparency records, neither of which existed during the previous assessment.

[53] See Managing the Risks of Top-Level Domain Name Collisions: Findings for the Name Collision Analysis Project (NCAP) Study 1, https://www.icann.org/en/system/files/files/ncap-study-1-report-19jun20-en.pdf

[54] See Approved Board Resolutions | Regular Meeting of the ICANN Board | 25 March 2021, https://www.icann.org/en/board-activities-and-meetings/materials/approved-resolutions-regular-meeting-of-the-icann-board-25-03-2021-en#2.b

[55] See SSAC 2021-02: Revised Study Two Proposal for the Name Collision Analysis Project, https://itp.cdn.icann.org/en/files/security-and-stability-advisory-committee-ssac-correspondence/ssac2021-02-05feb21-en.pdf

2. General DNS Evolution and Observational Impairments: DNS usage monitoring provides insight into time-resolved traffic evolution patterns useful in the quantification of system stability and performance as well as detecting aberrant events. Longitudinal measurements and usage trends, however, are increasingly difficult to leverage as the underlying system evolves or as bifurcation within the system occurs. These system changes may result in non-symmetric system usage, partial or even total impairments in DNS measurements, and ultimately confound the interpretability of the system's usage metrics. Since the last round of TLD delegations, several new technologies and recommended best practices within the DNS ecosystem now have a significant impact on the volume and fidelity of DNS queries observed at nameservers in the DNS hierarchy. These technologies include running Root on Loopback (RFC 7706), Aggressive Use of DNSSEC-Validated Cache (RFC 8198), DNS Query Name Minimization (RFC 7816), and DNS Queries over HTTPS (RFC 8484). It is in the DNS community's best interest to develop a better understanding of how these standards and technology changes will influence data collection capabilities as well as their impacts to data analysis of DNS traffic in an ever evolving, technologically fragmented, and highly distributed system.

3. Controlled Interruption Efficacy and Data Analysis: While the NCAP Study One Report highlights some anecdotal reports around the efficacy of Controlled Interruption, a thorough assessment of the framework has yet to be started. The collected reports should at a minimum be analyzed to better understand any trends, commonalities, faulty assumptions, and success attributes. Understanding the nature of these reports with a re-examination of previous DITL data may help identify key signals in the DNS that could better inform name collision risk assessments moving forward. Some applications, including popular browsers, have implemented specific DNS controls to signal when Controlled Interruption events occur. To that end, efforts should be made to identify and contact such vendors to see if instrumentation data is available. Finally, a study should be made to provide evidence that Controlled Interruption was a successful mitigation model, which may include creating and running simulation test beds.

4. Vulnerability Understanding and Mitigation Strategies: Since the last delegation of TLDs, various peer reviewed academic and industry papers have been published that elucidate some of the more detailed nuances of name collisions, specifically as they relate to various risks and vulnerabilities. Specifically, many of these publications directly identify known DNS query patterns, typically associated with zero-configuration protocols such as DNS-SD, that can be weaponized and exploited in a name collision environment. This new knowledge should be applied to future TLD delegation risk assessments as it builds upon a foundational understanding of the intent of the DNS queries as opposed to the volume of queries that was originally used in the new gTLD risk assessment.

3.2 Review of Available Datasets

As part of the effort to build a workflow for evaluating name collision risk, the DG explored what DNS data is available for review. In addition to the DITL data and information from two recursive resolvers discussed in the perspective study, two additional areas were explored as possible sources for developing the necessary CDMs to evaluate name collision risk: Identifier Technology Health Indicators (ITHI) metrics and ICANN Managed Root Server (IMRS) DNS Magnitude data.[56]

On 4 August 2021, Alain Durand from the Office of the Chief Technology Officer at ICANN and Christian Huitema from Private Octopus presented[57] to the DG the ITHI project (started in 2017) monitoring the health of the registered identifiers ecosystem, through a set of ITHI metrics. There are eight detailed metrics for which data can be seen on the site dedicated to the ITHI project.

The metrics are computed using data captured from various sources including data collected by ICANN projects and traces obtained from participating root DNS servers, authoritative DNS servers, and recursive DNS resolvers. Recently, ICANN's Office of the Chief Technology Office has published the OCTO-25 document regarding the ITHI project[58], which includes an entire section dedicated to name collisions.

In addition to the ITHI data, the DG considered the data available from the ICANN Managed Root Server (IMRS) during the 3 November 2021 DG call (23:36 in the recording[59]), specifically as part of the ICANN DNS Magnitude project.[60] The ICANN DNS Magnitude project assumes that the number of unique networks that send DNS requests reflects the overall popularity of the domain's services. This DNS-based metric "DNS Magnitude" can be used for estimating the popularity of a domain. As per their website, they apply this ranking and classify top-level domains by their delegation status, and offer the advice that non-existent domains that are heavily queried for by a large number of networks have a high collision risk.

Both datasets are noted as possible sources of information that the Technical Review Team (TRT) (See Appendix 3) might use for information prior to root delegation.

The NCAP DG deliberated on the proposed data collection methods as a sample of possible and available methods based upon careful consideration and balance of data privacy risks and potential benefits.

The data collection methods proposed for the TRT are a small sampling of known and tested methods. Other methods may be used, but they remain untested within this report. Ultimately,

[56] https://www.icann.org/ithi-faqs and https://magnitude.research.icann.org/

[57] https://community.icann.org/pages/viewpage.action?pageId=169443849

[58] See Identifier Technologies Health Indicators (ITHI) Retrospective and Proposal, https://www.icann.org/en/system/files/files/octo-025-08jul21-en.pdf

[59] See NCAP Discussion Group - Weekly Teleconference, 3 November 2021, https://icann.zoom.us/rec/play/q_sQBiDJFQmNLxrala7bGNd2zHBCpLgxQbMndTbdj6FFAXjO2JLHN8VqUzO0yHGgBFGAa_-6Gte-itfk.gVQmFPkJCDlZ5l4i?continueMode=true&_x_zm_rtaid=-ogRgxjzQjuYlgz7OP-hWg.1659537978260.87e6fbcb4027d9be5b84c717c5fde600&_x_zm_rhtaid=833

[60] Also covered in a session held at ICANN 72 (https://72.schedule.icann.org/meetings/EpPBA8MefE5dw6Ymm)

which data sources and data collection methods to use should be critically considered during the operationalization of the TRT.

3.3 The Issue of Manipulation

One area of concern for the DG involves third-party manipulation of the CDMs used to evaluate the risks associated with name collisions. Discussed during ICANN 74 and on the 25 May 2022 call, there are a variety of ways a third party could fabricate the appearance of name collisions in the DNS RSI and resolver logs. At this time, there is no way to predict or prevent this type of manipulation, and identifying the data to differentiate between legitimate name collisions and fabricated ones requires longitudinal data analysis by the TRT.

Moreover, a determined attacker with enough lead time could hide the manipulation such that it would be challenging for the TRT to identify it. There is also a risk here that with the knowledge that the TRT, prospective registrants, or other parties will use the manipulated data creates an unintended incentive for this manipulation, which could result in very large numbers of unnecessary CDM queries, and thus requiring investigation that might delay Name Collision Occurrence Assessment by the TRT.

The DG agreed that reviewing the data and making this judgment call must be part of the responsibilities for the TRT. This is a difficult problem that will likely require unique, customized data analysis efforts that may or may not succeed in identifying manipulation. The issue of manipulation is a residual risk that must be accounted for by TRT analysis. This is discussed further in Section 4.3.

3.4 Critical Diagnostic Measurements

As highlighted in the Case Study of Collision Strings (hereinafter referred to as “Case Study”), recommendations regarding any course of action in handling name collisions is based on a set of CDMs and no single class of measurement is sufficient to assess the full scale of name collision risks.[61] The different measurements must be taken as a whole to understand how their interactions inform any technical analysis. For example, as described in the Case Study:

> query volume--one of the four [4] major classes of measurements--is an important factor, but a single source that could be easily mitigated with a simple configuration [change] may be responsible for high query of a name. Conversely, if not only query volume was high, but query origin diversity (i.e., from many networks and many systems) and query type diversity were also extremely high, this would suggest collision impact may be greater. This is because the expectation of negative responses is high, and the mitigation across multiple services, networks, and users is increasingly complex to perform.”

[61] See Case Study of Collision Strings, https://www.icann.org/en/system/files/files/case-study-collision-strings-13jul22-en.pdf

The four (4) major classes of measurement that should help assess the scope, impact, and potential harm of name collisions include, in no particular order:

- Query Volume – The number of queries each RSI receives
- Query Origin Diversity – The number of unique query source IP addresses (resolvers)
- Query Type Diversity – The type of query (i.e., resource record type) being requested
- Label Diversity – Diversity of labels under a Collision String

Along with these four (4) major classes of measurement, other characteristics identified as Critical Diagnostic Measurements include[62]:

- Open-Source Intelligence (OSINT)
- Qualitative assessments

Additionally, the Root Cause Analysis report introduced as an additional metric the number of unique DNS "suffixes" identified. These suffixes are DNS domains used by organizations to qualify otherwise unqualified DNS lookups being made from within.

These diagnostic measurements were among those used previously by JAS and Interisle to better understand and assess the risk of Collision Strings. As stated in the JAS Final Report, their taxonomy of name Collision Strings depended on:

> (1) the diversity of querying source IP addresses and Autonomous Systems; (2) the diversity of labels queried; (3) applying sophisticated 'randomness detection' to strings and substrings; (4) presence of linguistic terms and colloquialisms in strings and substrings; (5) temporal patterns; and (6) analysis of the Regular Expressions of the labels queried within each TLD and across all TLDs.

However, as previously discussed, the quality and availability of data to qualitatively or quantitatively assess name collisions is a significant and growing concern.

3.5 Generating Data for Evaluation

There are several potential methods for collecting data to evaluate the risk of name collisions. The different methods bring to light different CDMs and introduce new opportunities and risks through the data collected. The DG took into consideration concerns regarding privacy, potential user reactions, and application design when handling different notification signals, protocols, and architectures.

[62] Open-Source Intelligence (OSINT) and qualitative assessments are mentioned in the Case Study as other characteristics but for those strings that require a qualitative rather than a quantitative assessment. OSINT strings require research to understand the semantic meaning of the string and what that string could be associated with.

In the 2012 new gTLD round, there were static data sets and root server logs already in existence that served the purpose of providing a broad picture of DNS activity. Those data are no longer sufficient given the changes in DNS architecture over the last decade. With the introduction and widespread use of public resolvers, new methods to understand when and where name collisions are happening are required.

The Study Two DG ultimately came to consensus around the following four (4) methods of measurement to assess risk in relation to applied-for strings. All of the methods subsequently described involve delegating the applied-for string to servers managed by some entity, in conjunction with the name collision assessment process for that string. Besides the benefit of an Emergency Back-End Registry Operator (EBERO) not being needed since ICANN is in control of string delegation during the assessment process that precedes granting of a TLD, there are several other benefits to this delegation, as opposed to using root server query data, such as the day-in-the-life (DITL) data provided by DNS-OARC.

First, the set of authoritative DNS servers to which the applied-for string is delegated includes only related queries, as opposed to all queries that are received by root servers. Thus, the data set is less noisy. This achieves a similar effect as the trial delegation DNS Infrastructure Testing described in SAC062 as a mitigation strategy for name collision risks, where "the only names permitted to exist in the zone would be those required as part of the data collection or testing."[63]

Second, the servers and data are managed by a single entity, rather than a consortium of organizations. Whereas DITL provides a data set once per year, and not all root server operators fully participate, this consolidated management facilitates getting a more comprehensive, consistent data set in real-time. This tactic aligns with the mitigation measure of making "available to the single entity that is the sole originator of name collisions for that [TLD]" proposed within the Collision Occurrence Assessment described in the New gTLD Collision Occurrence Management Proposal.

Finally, by having control of the time-to-live (TTL) values for the records in the DNS zone associated with the applied-for string, the effects of caching can be mitigated, such that observed query volume more accurately reflects that of clients behind recursive resolvers. This action is informed by SAC062 as a benefit of "trial delegation" allowing for emergency rollback if any significant consequences occur.

3.5.1 "No Interruption" – DNS NODATA Response

The least intrusive method for collecting name collision data involves configuring servers authoritative for the applied-for string to return NODATA responses in response to queries for subdomains of that string. A NODATA response is an indicator that "the name is valid, for the

[63] See SAC062: SSAC Advisory Concerning the Mitigation of Name Collision Risk, https://itp.cdn.icann.org/en/files/security-and-stability-advisory-committee-ssac-reports/sac-062-en.pdf

given class, but [there] are no records of the given type"[64] (see Figure 2). It represents a change in behavior from the NXDOMAIN (name error) response that is issued prior to the delegation of the applied-for string. However, applications that originate such queries are not expected to behave differently with the NODATA response; thus, no disruption is anticipated (i.e., "no interruption") to be experienced by a user.[65] With NODATA responses, resolvers are forced to use the full query name in a DNS query, where it might not otherwise be included, due to negative caching and QNAME minimization. This increases and enriches the data available to analysts for assessing potential name collision issues associated with the applied-for string.

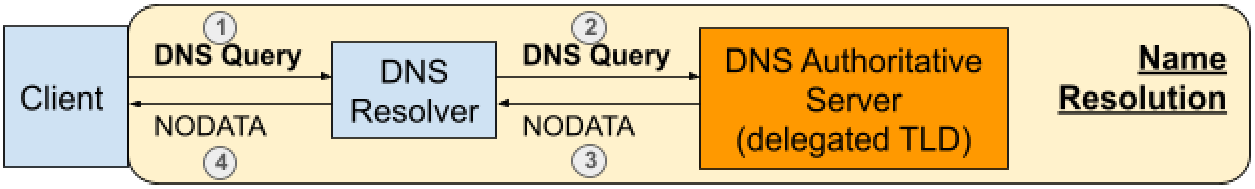


Figure 2: Representation of DNS NODATA response ("No Interruption")

Implementation. The DNS NODATA responses are accomplished by limiting the zone contents for the delegated string to only 1) requisite SOA and NS records for the zone itself and 2) a wildcard record of type HINFO[66]. Queries for type HINFO have no meaning and thus are not anticipated. Queries for anything other than HINFO will always result in NODATA responses, i.e., a NOERROR response code but no answer data. The time-to-live (TTL) value and "minimum" SOA field are set to a value of 60 seconds, to minimize the effects of negative caching[67]. A full example of a zone using this configuration is shown in Appendix 2.

Logging. With this method, all DNS queries associated with the applied-for string are logged. Among the features logged are: timestamp, client IP address, client port, server IP address, server port, IP version, transport-layer protocol, query name, and query type.

3.5.2 "Controlled Interruption" – Transport-Layer Rejection at Local System

The purpose of the method described in the previous section (i.e., "no interruption") is to collect name collision data with minimal disruption to end-users or -systems. However, that method provides no mechanism for informing end-users and -systems that they are experiencing name collisions, in the hope that such notification will elicit a configuration change. This next method introduces an intentional disruption to provide one type of notification.

[64] See RFC 2308: Negative Caching of DNS Queries (DNS NCACHE), https://www.rfc-editor.org/info/rfc2308

[65] Members of the NCAP DG performed extensive testing of library and application behavior where NODATA responses were returned instead of the NXDOMAIN responses returned prior to delegation of the TLD string. See implementation experience in RFC 8482, Section 8 https://www.rfc-editor.org/info/rfc8482.

[66] Similar methodology has been specified for responding minimally to responses of type ANY. See RFC 8482: Providing Minimal-Sized Responses to DNS Queries That Have QTYPE=ANY, https://www.rfc-editor.org/info/rfc8482

[67] See RFC 2308: Negative Caching of DNS Queries (DNS NCACHE), https://www.rfc-editor.org/info/rfc2308

This is done by configuring servers authoritative for the applied-for string to return a specific IPv4 address in response to queries – an IPv4 address that is routed to and only usable by the local system itself. The very presence of this IP address prompts applications using a collision name to initiate communication with that IP address. That communication is directed only to the local system and thus not observed on the Internet. However, the local system is almost certainly not expecting that communication, so the communication is rejected or simply ignored at the transport layer (see Figure 3). Affected applications are expected to fail with a message and behavior that depends on the application. Despite the intentional and inevitable disruptions encountered by users and systems experiencing controlled interruption, the hope is that those disruptions prompt affected parties to investigate and fix the problem. Without such remediation of these artificial collisions early on in the delegation process, affected users and systems run the risk of encountering name collisions with some other third party that has registered the name, with potentially more dire consequences.

Controlled interruption was the method exclusively used in the 2012 round of applications. As this method has a deployment history, some amount of analysis has been done on controlled interruption, including user and system impact, root cause discovery, and overall effect on DNS queries associated with the string. These analyses include the NCAP Study 1 Report and the Root Cause Analysis. According to those reports, the level of impact on users and systems disrupted ranged from negligible impact to significant impact. There is significant evidence from Web searches that the controlled interruption IP address was discovered and asked about in online forums. However, a minority of surveyed users that were affected discovered the IP address or found it helpful in identifying the cause of their problems.

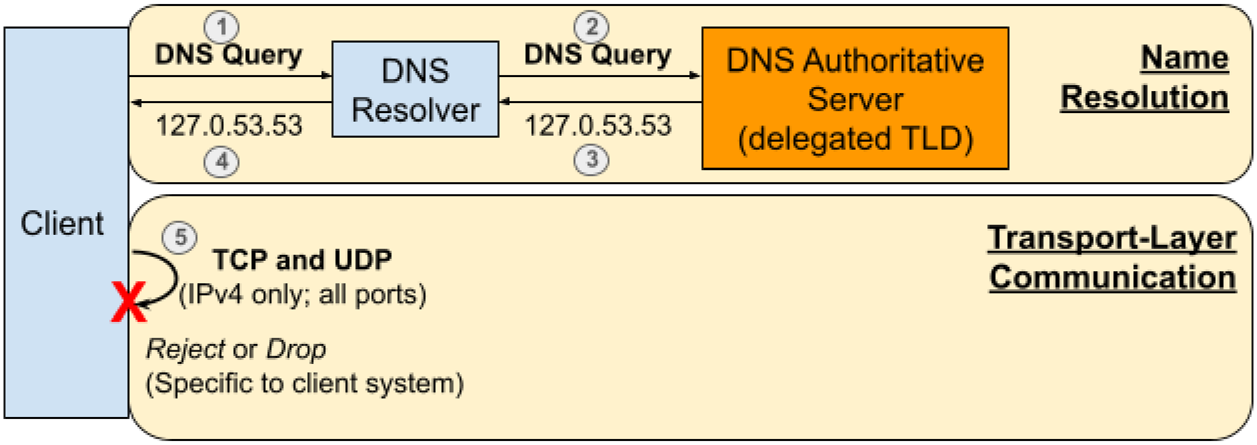


Figure 3: Representation of Transport-Layer Rejection at Local System (“Controlled Interruption”)

The DG notes that there is no exact IPv6 equivalent of the IPv4 addresses used for controlled interruption. While IPv6 solutions have been mentioned in DG meetings, none have been thoroughly discussed or tested. IPv6 is a risk tradeoff which was thoroughly discussed in the JAS report. There is no clear, risk-free approach to 2012-style CI in v6 space. For this reason, controlled interruption, as proposed, only works with IPv4. Despite this apparent shortcoming, this only affects notification for the few, if any, affected hosts that have IPv6-only connectivity.

Implementation. With controlled interruption, the zone is configured in the same manner as the previous section (“no interruption”), but additionally the zone contains wildcard records of type A (IPv4 address), MX, and TXT. The record data for each of these types is composed of values that prevent an application from initiating transport- or application-layer communications outside of its own system. The IP address returned in response to queries of type A is 127.0.53.53, which is within an IP block for which communication is never routed outside a local system.

Logging. With this method, all DNS queries associated with the applied-for string are logged in the same way as with the “no interruption” method.

3.5.3 “Visible Interruption” – Transport-Layer Rejection at Public IP

The method described in the previous section (“controlled interruption”) adds a mechanism for potentially interrupting applications in an effort to notify them of the potential name collision problem. However, because transport-layer communications never leave the local system with that method, the interruptions cannot be observed by any external entity. To address that deficiency, this method makes the data associated with these interruptions available for analysis by doing the following. Authoritative DNS servers are configured to return an IP address, but this time the IP address corresponds to a server on the Internet, managed by an entity involved in assessing name collisions. This “sinkhole” server is configured to cause the same interruption behavior observed with controlled interruption (See Figure 4). Thus, end-user and -system application behavior is interrupted, but attempts to communicate with the IP address are routed outside the local system to the sinkhole server, where they can be used for analysis (i.e., “visible interruption”).

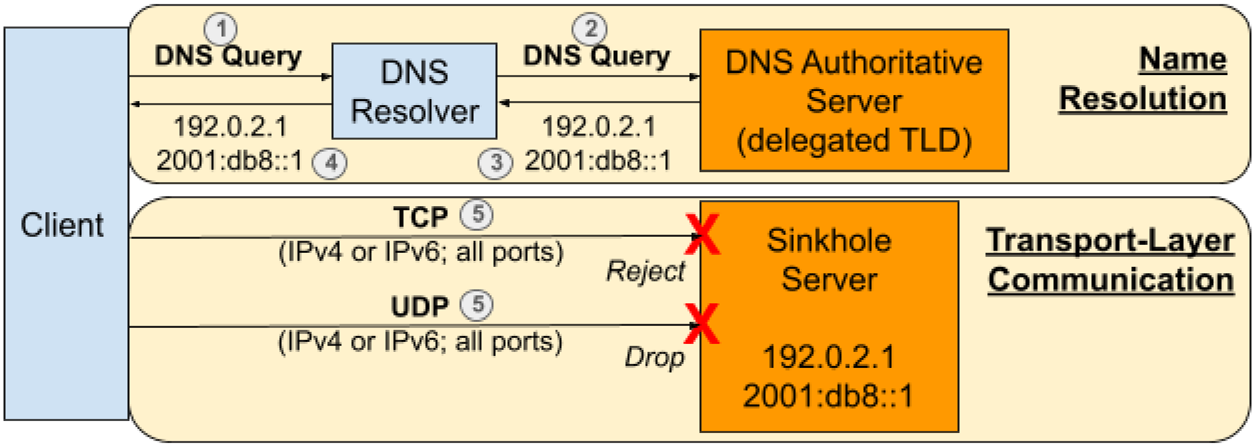


Figure 4: Representation of Transport-Layer Rejection at Public IP (“Visible Interruption”)

Implementation. With visible interruption, the DNS zone is configured in the same manner as it is with controlled interruption, with the following differences. The IPv4 address associated with the wildcard A record corresponds to the sinkhole server. Additionally, a wildcard AAAA record is introduced into the zone with an IPv6 address that corresponds to the sinkhole server. Reverse DNS entries for the IPv4 and IPv6 addresses (i.e., within the in-addr.arpa and ip6.arpa domains) map to PTR records that provide a meaningful message encoded into a domain name. Additionally, it would be desirable for the IPv4 and IPv6 themselves to be meaningful and recognizable, just as the controlled interruption IP address (127.0.53.53) has been. The sinkhole server is configured to actively reject incoming TCP connections and ignore incoming UDP datagrams.

Logging. With this method, all DNS queries associated with the applied-for string are logged in the same way as with the “no interruption” method. Additionally, at the sinkhole server, all communication attempts are logged. Among the features logged are: timestamp, client IP address, client port, server IP address, server port, IP version, transport-layer protocol, and TCP header values (TCP only).

3.5.4 “Visible Interruption and Notification” – Transport-Layer Rejection and Application-Layer Notification at Public IP

The methods described in the previous two sections (“controlled interruption” and “visible interruption”) use application disruption as a means to communicate to the end-user or -system that they are experiencing name collisions. Both methods leave hints to the affected parties as to the cause of the problem. However, neither method directly and explicitly informs the user of the problem. The next method follows the pattern of visible interruption, but instead of universally rejecting incoming communications on all ports, the sinkhole server is configured to accept application-layer communications for a small subset of ports and services and to return a descriptive response of the name collision problem via the corresponding protocol (See Figure 5). The only proposed protocol is HTTP on port 80.

Other ports and protocols were considered, including HTTPS on port 443, but because of unresolved challenges with technical implementation and/or end-user experience, only HTTP was left as an option. Thus, end-user and end-system application behavior is interrupted, and communication attempts are visible at the sinkhole server. However, browsers communicating with HTTP on port 80 will receive a notice about the name collisions that can potentially be processed by the end-user or -system (i.e., “Visible Interruption and Notification”).

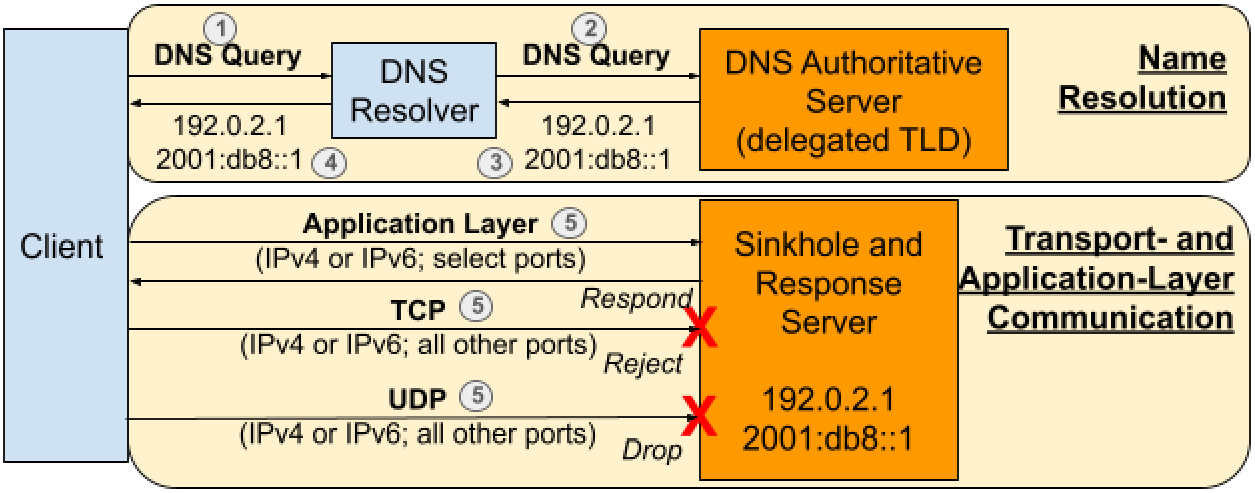

Client
1
DNS Query
DNS Resolver
2
DNS Query
DNS Authoritative Server (delegated TLD)
Name Resolution
192.0.2.1
2001:db8::1
4
3
192.0.2.1
2001:db8::1
Application Layer
5
(IPv4 or IPv6; select ports)
Sinkhole and Response Server
Transport- and Application-Layer Communication
TCP
5
Respond
(IPv4 or IPv6; all other ports)
Reject
UDP
5
(IPv4 or IPv6; all other ports)
Drop
192.0.2.1
2001:db8::1

Figure 5: Transport-Layer Rejection and Application-Layer Notification at Public IP ("Visible Interruption and Notification")

Implementation. With visible interruption and notification, the DNS zone and sinkhole server are configured in the same manner as they are with "visible interruption", with the following differences. Instead of rejecting TCP communications to port 80, the sinkhole server runs an HTTP server on port 80 that responds to all incoming HTTP requests with a 302 Redirect HTTP response code. This response directs the HTTP client to a page with more information on name collision.

Logging. With this method, the logging of DNS queries and transport-layer communications are the same as with the "visible interruption" method.

3.6 Benefits, Potential Harms, and Privacy Considerations of Proposed Methods

The DG recognizes that there are both perceived benefits and potential harms associated with each one of the proposed methods. The specifics of each method are summarized in the following table, which is explained hereafter.

Method	**Disruption**	**Notification**	**History**	**Privacy / Telemetry** **D= DNS Recursive-to-Authoritative Queries** **T= Transport-Layer Communication Attempts** **A= Application-Layer Data** **H= HTTP Request, OS, Browser/Client Version**	
				Disclosed	**Logged**
No Interruption	No	None	None	D	D
Controlled Interruption	Yes	1. Transport-layer disruption; 2. Domain names resolve to 127.0.53.53, which can be searched for on the Web.	2014 - present	D	D
Visible Interruption	Yes	1. Transport-layer disruption; 2. Domain names have meaningful reverse DNS entries that refer to ICANN.	None	D, T	D, T
Visible Interruption and Notification	Yes	1. Transport-layer disruption; 2. HTTP server returns a special response to direct clients to information on name collisions. 3. Domain names have meaningful reverse DNS entries that refer to ICANN.	None	D, T, A, H	D, T

Disruption is both a benefit and a potential harm. The benefit is that it is an avenue for notification. The harm is that it potentially disrupts applications of users and systems using the affected string, sometimes at large scale. There are examples of this in the Root Cause Analysis. Only the no interruption method is expected to avoid disruption altogether.

Notification is a benefit associated with all methods except no interruption. The controlled interruption and visible interruption methods attempt to notify by both disrupting application behavior and leaving a hint as to the cause of the disruption. The visible interruption and notification method attempts to notify by providing a human-readable message.

Only controlled interruption has any **history** of deployment, as it was the only method used during the 2012 round. While controlled interruption has both pros and cons (noted in section 3.5.2), the fact that it is the only method with a history makes it stand alone in that regard.

Information disclosure concerns potential **privacy** harms. In this case, users or systems potentially send to the public Internet information that would likely not have been exposed otherwise. In the case of no interruption and controlled interruption, only DNS queries are leaked, many of which would already be observable on the public Internet. However, with visible interruption, transport-layer data is also shared on the public Internet. With visible interruption and notification, application-layer data is shared on the public Internet. With regard to logging, relevant telemetry data is stored for analysis, except for application-layer data.

The DG noted concerns about the use of these methods of data gathering, mostly but not entirely around the privacy of users or organizations who can't feasibly be informed or asked for consent regarding data collection on public infrastructure. Discussion on this topic included both ethical considerations and associated legal or reputational risk, such as the potential for negative publicity or liability under privacy laws.

The DG had broad consensus that the methods proposed provided the most viable options to support future assessments. Additionally, the DG at large recognized that there are benefits in each of the proposed tools, along with potential privacy risks associated with the use of some.

The DG also widely agreed that this report should document the techniques we know about for collection and use of name collision related data, draw awareness to potential risks associated with these tools, and leave assessment of when their use is appropriate or the sequence of methods for assessment to the Technical Review Team.

Privacy and legal risks related to the use of data collection methods are not new to this study as the fine balance between identification and notification of potential name collisions with privacy protection involves the exercise of judgment.[68]

[68] See ICANN Name Collision Occurrence Management Framework, ICANN, https://www.icann.org/en/system/files/files/name-collision-framework-30jul14-en.pdf

As the DG is neither in a position to assess such non-technical risks nor to operationalize mitigation strategies, the DG looks to the ICANN org to implement the relevant recommendations and necessary procedures required to limit potential negative impacts to the DNS and the ICANN org, including a data privacy and protection policy, along with other appropriate risk mitigation measures for legal compliance.

4 Findings

After reviewing years of earlier work, including the Interisle Report[69], the JAS Report[70], and the NCAP Study 1 Report, as well as the outputs of the three studies included as part of the NCAP Study 2 efforts (the Case Study of Collision Strings, the Perspective Study of DNS Queries for Non-Existent Top-Level Domains, and the Root Cause Analysis ("RCA") reports), the NCAP Discussion Group (NCAP DG) made several observations regarding the issues surrounding name collisions.[71] These findings ultimately informed the recommendations offered by the NCAP DG later in this report.

4.1 The definition of what is a name collision has evolved over time

Recommendation 2 - ICANN should adopt a consistent definition for name collision

Section 1.2, "Background and Related Work," reviews the history of defining a name collision, ending with the definition developed by the NCAP DG and used to scope the work of NCAP Study 1 and Study 2 (Section 1.1), repeated here for convenience:

> Name Collision refers to the situation where a name that is defined and used in one namespace may also appear in another. Users and applications intending to use a name in one namespace may actually use it in a different one, and unexpected behavior may result where the intended use of the name is not the same in both namespaces. The circumstances that lead to a name collision could be accidental or malicious. In the context of top-level domains (TLDs), the conflicting namespaces are the global Internet Domain Name System (DNS) namespace reflected in the root zone as published by the Root Zone Management Partners and any other namespace, regardless of whether that other namespace is intended for use with the DNS or any other protocol.

When considered in the scope of work for Study 2 (see Section 1.1), two important conclusions apply to the discussion group's work.

First, ICANN only has a role in managing one namespace: the global Internet Domain Name System. Thus, the scope of the analysis and recommendations of name collisions in this study is focused on identifying and mitigating name collisions with the global Internet DNS.

Second, identifying and mitigating name collisions exclusively within alternate naming systems is out of scope for the NCAP DG (see Appendix 1), which has focused on name collisions

[69] See Name Collision in the DNS ("Interisle Report"), https://www.icann.org/en/system/files/files/name-collision-02aug13-en.pdf

[70] See Mitigating the Risk of DNS Namespace Collisions: Final Report ("JAS Report"), https://www.icann.org/en/system/files/files/name-collision-mitigation-final-28oct15-en.pdf

[71] See Managing the Risks of Top-Level Domain Name Collisions: Findings for the Name Collision Analysis Project (NCAP) Study 1, https://www.icann.org/en/system/files/files/ncap-study-1-report-19jun20-en.pdf

between names in the public DNS and other namespaces (such as an organization's internal namespace or a non-DNS namespace.) However, the construct of a DNS fully qualified domain name (normally presented to a user as a sequence of labels separated by a ".", e.g., "www.icann.org") is being used in other namespaces. This usage confuses both users and the applications and services that users rely upon when navigating the Internet.

The analysis proposed in this study will result in many of these usages becoming visible and included in the metrics for identifying name collisions. However, it is out of this study's scope to seek to identify these name collisions and recommend mitigation for use in these other namespaces. Nonetheless, the proposed Technical Review Team, introduced in Appendix 3–"Collision Assessment Workflow Development"–should note the existence of other namespaces as they are discovered in the data.

4.2 Name Collision Identification and Quantification

Recommendation 1 - ICANN should treat name collisions as a risk management problem

Recommendation 5 - ICANN must support the delegation of strings in order to improve the ability to conduct a name collision risk assessment

Drawing from the Case Study of Collision Strings (see Section 2.1 and the associated report) and the Root Cause Analysis (RCA) Reports (see Section 2.3 and the associated reports), there are no guarantees when it comes to identifying and mitigating name collisions. Quantitative data analysis only produces indicators of visible name collisions, but the questions of whether or not there is an actual name collision problem, how broad the population of affected users or systems is, and what level of harm is or would be experienced cannot be definitively answered without qualitative analysis. Understanding the implications, including the level of harm, depends on data beyond what is available in any aggregation of log files or historical data. As noted in the Case Study, "No one measurement alone is generally going to provide sufficient quantitative or qualitative indications to thoroughly assess the name collision risks expressed by a string."[72]

Potential indicators of impact and risk can be learned from the available data. To definitively ascertain the level of impact or even the existence of any particular name collision, any quantitative analysis must be combined with a qualitative assessment. Nonetheless, the RCA shows a positive correlation between quantitative and qualitative assessments of available data: "the name collision reports are supported strongly by measured data."[73] Even with that finding,

[72] See Case Study of Collision Strings, https://www.icann.org/en/system/files/files/case-study-collision-strings-13jul22-en.pdf
[73] See Root Cause Analysis - New gTLD Collisions, https://www.icann.org/en/system/files/files/root-cause-analysis-new-gtld-collisions-18jan23-en.pdf

the RCA suggests additional studies that include “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.”

During its discussions, the NCAP DG observed that there is a need in many, if not all, cases to apply human judgment when analyzing critical diagnostic measures (CDMs). While the NCAP DG agreed that having numerical definitions for “high” and “low” would make the initial evaluation of a name collision more straightforward, any attempt to come to that definition resulted in an intractable debate. The principal issues are presented below.

As noted in Finding 4.2.2, the Perspective Study of DNS Queries for Non-Existent Top-Level Domains shows that query data does not always reach the root servers. Query Name (QNAME) minimization (QNM), aggressive caching, and local resolver features increasingly affect the nature of queries seen at the root servers. For example, the volume and diversity of queries observed at the root servers were shown to be different from the volume and diversity of queries observed at at least one recursive resolver. Recursive resolvers (both globally public recursive resolvers and private enterprise recursive resolvers) are deploying solutions to locally manage their own known list of TLDs or minimize the amount of data and queries sent to authoritative servers. As long as access to query data is restricted—an action that may be done for good reason (e.g., to decrease latency or protect privacy)—name collisions will not always be visible.

In some cases, the issue involves internal name collisions within network systems. These name collisions are often undetectable when analyzing available data sets such as root server logs or Day-In-The-Life (DITL) data. However, even if the names are not leaking into the DNS, the issue of name collision still matters to the people using and the people applying for the name.[74]

Despite their invisibility to these external measurements, internal name collisions significantly impact network users and administrators. These collisions occur when different entities within the same network use identical identifiers, leading to confusion and potential system errors. This situation is particularly problematic for network users attempting to access specific resources, as well as for individuals applying for new names or identifiers within the system. The resolution of these collisions is crucial for maintaining efficient network operations and ensuring a seamless user experience.

Given the fact that not all name collisions can be made visible, there will always be some amount of risk with the technical delegation and applicant delegation granting of a new TLD string, regardless of whether it has evinced a name collision in the DNS telemetry data.

[74] See “Losing Visibility into Dns,” 25 February 2022, https://wkumari.github.io/2022/02/25/losing-visibility-into-dns.html

4.2.1 Name collisions continue to persist within the DNS

Recommendation 1 - ICANN should treat name collisions as a risk management problem

Recommendation 3 - ICANN should continue its education and outreach efforts to the community on the name-collision topic

Name collisions cannot be predicted or prevented with any consistent degree of certainty, and new instances of name collision, even for reserved TLDs, may happen at any time. As shown by the examples of .CORP, .HOME, and .MAIL, name collisions continue to occur even ten years after their original identification as Collision Strings. Additionally, as seen within the Root Cause Analysis (RCA) Report, a name collision scenario was enacted on a TLD string that was delegated nearly 15 years prior due to a network manufacturing company erroneously setting a default configuration value to “domain.name”.

Other examples of name collision growth and exacerbation due to pandemic conditions and transient devices being used in their non-corporate environment are evident in the heightened CDMs shown in the longitudinal analysis of .HOME and .CORP.[75] A logical conclusion based on this data is that name collisions are likely to persist in the DNS; new instances of name collision may happen at any time and thus any name collision assessment is a “point-in-time” analysis.

4.2.2 There are limitations with using currently available data sources for understanding root cause and risk, or designing mitigation and remediation plans

Recommendation 1 - ICANN should treat name collisions as a risk management problem

Recommendation 7 - ICANN should establish a dedicated Technical Review Team function

Recommendation 11 - ICANN should not move ahead with NCAP Study Three

Currently available data sources and measurement methods might be insufficient for understanding root cause and risk, or designing mitigation and remediation plans. Even with the existing data, there is uncertainty that requires reviewers to make decisions on a string-by-string basis. In order to retain transparency and credibility of these judgments, they need to be based on the best available data and analysis as part of a formal review process.

Additionally, the formal review process for strings should be subject to a typical technical evaluation process without preferential review treatment for any grouping of strings. The

[75] See Case Study of Collision Strings, https://www.icann.org/en/system/files/files/case-study-collision-strings-13jul22-en.pdf

implementation of special procedures for certain types of strings based upon policy adoption is out of scope for this report.

In the 2012 round of new gTLDs, the analysis and resulting risk management framework was based primarily on root server DNS query data and Day-In-The-Life (DITL) query data. This served its purpose for the time. However, as noted in the revised proposal for Study 2, several infrastructure changes have contributed to reduced query visibility at the root servers. Thus, the efficacy of basing future analyses exclusively on root server query data is increasingly questionable.

Considering available datasets, it is worth noting that different datasets (e.g., DITL, ITHI, root zone logs) have different time-based characteristics. Some provide a dataset once per year (e.g., DITL), while others provide data in real time (e.g., root zone logs). Both views are necessary, though possibly not sufficient, to evaluate the likelihood that any given set of CDMs is a result of data manipulation. DITL itself, while still a valuable source of data, is limited by issues of data minimization and inconsistent data anonymization on the part of the root servers.

Further considering the issues of what is available in existing datasets, several new technologies and recommended best practices within the DNS ecosystem now have the potential to significantly impact the volume and fidelity of DNS queries observed at name servers in the DNS hierarchy since the 2012 round of gTLD delegations.

Coming back to the available datasets, the Perspective Study determined that they are often restricted by the non-standardized use of data anonymization techniques.

The Perspective Study of DNS Queries for Non-Existent Top-Level Domains (Section 2.2) shows that an analysis of query data from any proper subset of root servers will exclude query data from some fraction of Internet resolvers. Although a minimum number of queries from the majority of Internet resolvers will be seen, the report notes that caching and local resolver features affect the nature of queries seen at the root servers: the volume and diversity of queries observed at the root servers were shown to be different from the volume and diversity of queries observed at least one recursive resolver.

In addition to the decentralization of queries, the Perspective Study of DNS Queries for Non-Existent Top-Level Domains also shows that queries for a non-existent domain (NXDOMAIN) are increasingly less visible to root server operators as recursive resolvers (both globally public recursive resolvers and private enterprise recursive resolvers) deploy solutions to locally manage their own known list of TLDs or minimize the amount of data and queries sent to authoritative servers. The operational benefit to a recursive resolver of this type of solution is to reduce the latency of a class of queries by at least one transaction (a query and response with a root server), so it is understood why they would do this.

When queries for a potential Top-Level Domain (TLD) return a 'Non-Existent Domain' (NXDOMAIN) response, it becomes evident that a name collision is likely to occur for that TLD. This suggests that one way to measure actual harm would be to investigate the source of every NXDOMAIN query and evaluate if it would be harmful for that query transaction to fundamentally change. However, as an engineering reality, this is impractical, in part because of the volume that would need to be investigated and in part because of the ephemeral method with which IP addresses can be assigned.

Existing systems or name collision data repositories, such as ITHI and the ICANN DNS Magnitude Page, can provide some level of initial indication of a string's potential name collision impact. Current measurements from the ICANN DNS Magnitude Page show the large Pareto distribution of CDMs for the top 2,000 strings observed at ICANN's IMRS, in which there is nearly five orders of magnitude difference from the most queried string .INTERNAL with 288M queries per day and the lowest .HYPEMARK1 with 3.3K queries per day.[76] This data can only assist with providing a leading indicator of potential impact. Determining the harm solely from CDMs is unachievable. It is also worth noting that without sufficient longitudinal name collision data baselines, the manipulation of CDMs is problematic and again highlights the problematic nature of using CDMs to determine the potential of harm.

4.2.3 .CORP and .HOME demonstrated that high volume is an insufficient measure for analyzing the potential of high-risk impact

Recommendation 4 - ICANN should consider the need for mitigation and remediation efforts for high-risk strings

Recommendation 4.1 - ICANN should submit .CORP, .HOME, and .MAIL through the Name Collision Risk Assessment Process

Recommendation 8.1 - ICANN should not reject a TLD solely based on the volume of name collisions

In the "Case Study of Collision Strings," a method of identifying the impact of name collisions was developed, i.e., the impact of a name collision is based on both the volume of the queries and the diversity of the queries. The purpose of both is to identify the size of the parties affected by the collision and the potential for remediation of the collision. This is not an exact science.

Reviewing .HOME, the string with the most NXDOMAIN queries from the 2012 round, the NCAP DG observed a high volume of DNS queries that continues to increase and a significant diversity in the source of the queries. Equally important when considering the diversity of the source is that there is no discernable pattern to suggest that a single or small number of services

[76] See "Welcome to the ICANN DNS Magnitude statistics page," ICANN, accessed 19 December 2023, https://magnitude.research.icann.org

or applications are generating those queries. This could be considered in the 2012 round in part because DNS labels beyond just the TLD label in a query were visible; this information is increasingly less visible as various privacy-enhancing mechanisms are deployed in the DNS infrastructure.

Reviewing .CORP, the NCAP DG observed a string with significant NXDOMAIN queries from the 2012 round and a high volume of DNS queries that continues to increase with an apparent concomitant increase in the diversity of the source of the queries. In this case, investigation suggests that the principal cause of these queries is a globally dominant software package.

On the one hand, it is clear that the impact of both of these cases is high risk as there is a large number of globally dispersed users (including application clients) that would be affected by a change in the DNS behavior if the TLD string were to be delegated. This could intuitively suggest that there is an increased probability of harm, but it is difficult to know this with any certainty without additional data.

On the other hand, these two TLD strings have different diversity characteristics. In the case of .HOME, there is no discernible pattern to the globally diverse source of the queries, nor was there any single dominant source identified during the investigation. In contrast, the investigation of .CORP was able to identify a dominant cause for the source of the queries: Microsoft products that used “corp” as a default configuration option.[77]

Different CDM characteristics will have different implications when assessing risk. A high CDM does not definitively affirm high risk, nor does a low CDM imply low risk; this is why qualitative review is necessary.[78]

4.2.4 It is possible that future name collisions may occur on the scale of .CORP, .HOME, and .MAIL

Recommendation 8.2 - ICANN should request special attention to strings with high-impact risks during the name collision assessment process

As noted above, name collisions continue to persist in the DNS; it is reasonable to expect they will continue far into the future. Working with that expectation, it is worth noting that there may be additional Collision Strings on the scale of .CORP, .HOME, and .MAIL. As an example, the Case Study of Collision Strings identified six strings that met the early thresholds set by .CORP, .HOME, and .MAIL.

[77] From the JAS Report: “Many – but not all – queries seem related to Microsoft Active Directory systems which very often are rooted in “.CORP” per an unfortunate Microsoft configuration example more than a decade ago.”

[78] Verisign has done some work showing that remediation can be successful when a dominant cause for excessive, non-productive traffic can be identified, investigated, and resolved with the source. See “Verisign Outreach Program Remediates Billions of Name Collision Queries,” Verisign blog, 15 January 2021, https://blog.verisign.com/domain-names/verisign-outreach-program-remediates-billions-of-name-collision-queries/.

“As for the strings to be studied, the NCAP Revised Proposal asked for case studies of CORP, MAIL, HOME, and non-delegated strings that receive more than 100 million queries per day at the root. Using this threshold and DNS query data from A and J root servers results in six strings:.CORP, .HOME, .INTERNAL, .LAN, .LOCAL, and .MAIL.”[79]

Understanding that large-scale name collisions are a potential risk for delegated and un-delegated strings is a necessary part of the risk assessment for name collisions. Predicting when these large-scale collisions might occur is not possible.

4.2.5 It is impractical to create a do-not-apply list of strings in advance of new requests for delegation

Recommendation 9 - ICANN should create a Collision String List

Recommendation 9.1 - ICANN should support a mechanism that allows applicants to request a string be removed from the Collision String List

Because real-time quantitative and qualitative analysis is necessary to conduct a name collision risk assessment, it is impractical to create a "do-not-apply" list in advance. Any such list is subject to changes outside of ICANN’s control. Quantitative data is available that allows limited inferences, but qualitative data is also necessary to help validate the quantitative data; analysts must rely on the data to determine if a string is likely to be subject to a name collision that is at significant risk for causing harm.

4.2.6 Summary of Finding 4.2

Finding 4.2 underscores that quantitative data alone, such as logs and historical data, are insufficient to definitively assess name collision risks. This limitation is due to the inability of quantitative analysis to provide a complete picture of the extent and impact of name collisions. It is necessary to combine quantitative analysis with qualitative assessment to ascertain the true level of impact or even the existence of name collisions. This approach is supported by the positive correlation observed between quantitative and qualitative assessments in the RCA.

Furthermore, this finding points out the challenges in using current data sources for understanding root causes and risks or for designing mitigation and remediation plans. These challenges arise from changes in the DNS infrastructure, such as the growth of global public resolvers and the implementation of new technologies and practices that impact DNS query visibility. Not all name collisions are visible externally, such as those internal to a network, yet they remain significant for those using or applying for the name.

[79] See Case Study of Collision Strings, https://www.icann.org/en/system/files/files/case-study-collision-strings-13jul22-en.pdf

Regarding .CORP, .HOME, and .MAIL, high query volume is not a sufficient indicator of high-risk impact. The complexity and diversity of query sources further complicate the assessment of risk and impact. It is impractical to create a pre-emptive "do-not-apply" list for gTLD strings due to the dynamic nature of the DNS and the need for real-time, comprehensive analysis.

4.3 Data Manipulation Risks

The evolution of name collisions from accidental occurrences to potentially deliberate actions in future rounds is a significant concern. This shift necessitates a more rigorous analysis to determine the nature of these collisions. The findings in this section acknowledge that determining whether a collision is accidental or intentional is challenging, given the current technological limitations.

4.3.1 There is a risk of CDM data manipulation

Recommendation 7 - ICANN should establish a dedicated Technical Review Team function

Recommendation 6 - ICANN should establish and maintain a longitudinal DNS name collision repository in order to facilitate risk assessments and help identify potential data manipulation

As noted earlier in this document, there are a variety of ways a third party could fabricate the appearance of name collisions in the DNS root server instance and resolver logs. At this time, there is no way to predict or prevent this type of manipulation. Identifying the data to differentiate between legitimate name collisions and fabricated ones requires a level of review that offers flexibility and discretion as to what data to review and how to interpret that data.

To limit the potential manipulation of CDM measurements, reviewers may use longitudinal and historical data as one input to discover aberrant changes. But even with such data available, reviewers may find that long-run manipulation efforts are undiscoverable in the baseline. Depending on the design of the next application round, there may be critical points within the application process that present opportune moments in which manipulation of CDMs could impact the name collision assessment process.

In the 2012 round, the issue of name collisions included an assumption that the existence of any name collision was accidental (e.g., individuals and organizations that made a mistake in configuration). In future rounds, there is a concern on the part of the NCAP DG that name collisions will become purposeful (e.g., individuals and organizations will simulate traffic with an intention to confuse or disrupt the delegation process).

Determining whether a name collision is accidental or purposeful will be a best-effort determination given the limits of current technologies.

4.3.2 Data manipulation has ramifications beyond the technical aspects of name collision that are influenced by when analysis occurs

Recommendation 7 - ICANN should establish a dedicated Technical Review Team function

Data manipulation has ramifications beyond the purely technical difficulties involved in identifying when it occurs. It may also impact the timing and quantity of legal objections issued against proposed allocations, how the coordination of the next gTLD round is designed, and contention sets and auctions.

Name collisions are now a well-defined and known area of concern for TLD applicants when compared to the 2012 round, which suggests that individuals and organizations looking to "game" the system are potentially more prepared to do so.

4.4 Quantitative and Qualitative Measurement Considerations

Effective measurement and interpretation of data communication are the primary two tenets of name collision management. As noted in the findings in this section, there are critical considerations when it comes to collecting the data, interpreting it, and suggesting actions. Absolute numbers do not provide sufficient information–they must be interpreted in context with other information–but even so, the data collection process can be improved.

In the 2012 round of new gTLDs, proposed new TLD strings were allocated and Controlled Interruption was put in place for those strings. All of the strings that went through Controlled Interruption remained allocated because no harm was observed. One influence on the timing and order of name collision analysis was that name collision risk was not originally accounted for in the 2012 New gTLD process. The NCAP DG feels that given the ICANN community knows more about name collision and its impacts now, name collision analysis should occur before allocating.

Although there is a risk with the delegation necessary to conduct data collection without prior investigation, the simplicity of this solution can not be understated. In addition, 10 years of experience suggests that no significant harm manifested from the 2012 round, albeit a limited number of ICANN name collision reports, even though most of the TLDs delegated had name collision risk. On the other hand, it is important to note that it only takes one name collision to cause significant harm, and given the wide variation in the volume of NXDOMAIN queries for strings in the 2012 round, a question to consider is what does the volume of queries really tell us?

The decision was made for the 2012 round to delegate and to review harm after the fact. No evidence or strategy has been identified to change the need to delegate in order to conduct data collection for analysis.

4.4.1 Critical Diagnostic Measurements are structurally quantitative and benefit from supplemental qualitative information

Recommendation 1 - ICANN should treat name collisions as a risk management problem.

Recommendation 8.1 - ICANN should not reject a TLD solely based on the volume of name collisions

Recommendation 8.2 - ICANN should request special attention to strings with high-impact risks during the name collision assessment process

Considering the cases of .CORP and .HOME, the NCAP DG saw that those TLDs consistently have unique characteristics in that their CDMs have magnitudes of difference from any other non-delegated strings. Identifying those clear outliers is simple. More generally, however, the NCAP DG determined that the quantitative measures available with CDMs must be balanced with qualitative information in order to determine the level of risk of name collisions and any associated harms.

With the "Case Study of Collision Strings," the research quantified the presence of name collisions by defining and applying CDMs, reinforcing the research described in the earlier JAS Report.[80] These CDMs were collectively used to assess name collisions from the perspective of the root servers, using both volume and diversity of queries, origins, query names (labels), and query types. The data shows that name collisions remain an issue even though the ICANN community is more than a decade past discovering their risk to the security and stability of the DNS. This suggests that name collisions will remain an issue for the foreseeable future and thus supports the continued need for risk management related to name collisions.

However, while the CDMs can be used to quantify the impact of name collisions on root server query traffic directly, they cannot more generally quantify the impact on end users or organizations without qualitative data. The report itself disclosed as a weakness its "inability to truly measure the harm that might manifest as a result of a delegation." Thus, the volume of query data provides a useful heuristic for considering the impact of name collisions, but analysts cannot expect the query data to produce an accurate assessment of impact by itself. It is possible for strings with relatively low CDM values to have a relatively high potential impact and strings with relatively high CDM values to have negligible potential impact. The NCAP DG expects that changes in the DNS ecosystem caused by the increased deployment of DNS technologies such as QNM and aggressive NSEC caching might make low CDM values an increasing reality—such that the correlation between CDM trends and impact might even be more prone to error.

[80] See Mitigating the Risk of DNS Namespace Collisions: Final Report ("JAS Report"), https://www.icann.org/en/system/files/files/name-collision-mitigation-final-28oct15-en.pdf

The use of CDMs within a name collision risk management framework can provide insights into the probability of impact, but additional qualitative data is required to deduce the severity of harm. The CDMs used in the 2012 round and further reaffirmed by the NCAP Discussion Group's research have shown that the volume of queries is not in and of itself an indicator of harm nor is diversity; however, these CDMs do provide a leading indicator as to the potential risk of impact to clients and the end user community. For example, the root cause analysis showed that where there were reports of problems (qualitative data), the CDMs were high (quantitative data). It is also worth noting that name collisions may not be observable or even manifest during the name collision assessment period.

4.4.2 The quantitative data in CDMs can be improved

Recommendation 5 - ICANN must support the delegation of strings in order to improve the ability to conduct a name collision risk assessment

Recommendation 8 - ICANN should replace the existing Name Collision Management Framework with the recommended Name Collision Risk Assessment Framework

While quantitative data does not tell the whole story when it comes to the risk of name collisions, it does provide necessary information. Improving the quality of data collected can be done using a variety of tools.

SAC 062, SSAC Advisory Concerning the Mitigation of Name Collision Risk, describes a few options for trial delegation. These options are broken down into two categories:

- DNS Infrastructure Testing (Type I)
- Application and Service Testing and Notification (Type II)

In terms of the benefits and risks to trial delegations, the additional data will allow for better decisions to be made. They also increase the risk of potential manipulation of the data. Finding the balance is part of the risk assessment process.

4.5 Notification to users of name collisions is a critical function and separate from assessment or remediation

Recommendation 3 - ICANN should continue its education and outreach efforts to the community on the name-collision topic

The NCAP DG extensively discussed a few unique retrospective observations of the 2012 round regarding name collision mitigation and remediation processes. One of the primary concerns was the sequencing in which name collision analysis, notification and outreach, and delegation actions were performed by ICANN. It was recognized then, and the DG believes now, that the opportunity to understand name collisions and reduce their impact was critical to ICANN’s good stewardship of the DNS, but there was limited opportunity to include notification to users and system administrators in the process ultimately used in 2012 to assess name collisions or the effectiveness of remediation.

Effective communication is critical when attempting to pass relevant information to impacted parties. Ideally, notification messaging is sent in a direct manner to the impacted parties with a priori knowledge that the target audience will consist of both technical system administrators and non-technical end-users.

The overall value of a name collision detection and alerting technique is based on several factors, including alerting effectiveness, impact on end-system operational continuity, security and privacy, user experience, root cause identification, anticipated public response, and telemetry. The three notification modalities – proactively communicating with potentially affected parties; notification via log files; and application-based errors – may work in some cases, but not in others. No single notification method is expected to be more effective than any other at notifying a user, whether that is a system administrator or an application end user, of a name collision.

4.5.1 Controlled Interruption as a notification method is effective in some but not all instances

The 2015 JAS Report provides a strong analysis of Controlled Interruption as a way to raise awareness among systems administrators, who were in turn encouraged to “proactively search their logs for this flag IP address as a possible indicator of problems.”[81] Similarly, focusing on the applications performing a DNS lookup that would expect an NXDOMAIN response would interrupt the action and potentially send an error to the user of that application.

Even before the JAS Report, the Interisle Report from 2013 noted that it “may be possible to identify the parties most likely to be affected by name collision, and to notify them before the proposed TLD is delegated as a new gTLD.”[82] Despite raising awareness of potential name collisions, the Interisle Report describes notification as possibly “ineffective without substantial concomitant technical and educational assistance” due to parties not understanding “potential risks and consequences of name collision and how to manage them.”

4.5.2 Other methods for notification may be used but remain untested.

Recommendation 3 - ICANN should continue its education and outreach efforts to the

[81] See Mitigating the Risk of DNS Namespace Collisions: Final Report (“JAS Report”), https://www.icann.org/en/system/files/files/name-collision-mitigation-final-28oct15-en.pdf

[82] See Name Collision in the DNS (“Interisle Report”), https://www.icann.org/en/system/files/files/name-collision-02aug13-en.pdf

community on the name-collision topic

Additional methods, beyond Controlled Interruption, for notification may be used. However, their feasibility, use, effectiveness, and impact remain untested. As we only have data for Controlled Interruption, we cannot make a sweeping statement that describes the impact of other notification methods for which we have no data.

This uncertainty about effective notification is a gap in handling of name collisions by affected parties. ICANN has conducted education and outreach efforts in the past, which have partially filled this gap. If there were known techniques that could be relied upon to provide both additional assessment or remediation of name collisions, and notification to users and system administrators, a separate education and outreach effort would no longer be necessary. However, there aren't, and it is.

4.5.3 The criteria for the use of ICANN's name collision reporting form negatively impacted its use

Recommendation 8.3 - ICANN should update its public-facing name collision reporting process

The RCA report includes an analysis of the name collision reports received by ICANN, as well as a more general assessment of name collisions. The name collision reports received were biased by the fact that the form explicitly invited only submissions by users experiencing an extreme level of harm (i.e., "If your system is suffering demonstrably severe harm as a consequence of name collision, please fill in the form below to report the incident.")[83]

The reporting form did not require contact information, and some individuals used it without meeting the expressed threshold. That said, the NCAP DG suspected that individuals were deterred from filling out the form, which limits what ICANN can learn from this mechanism. While requiring all individuals experiencing a name collision to fill out this form is unreasonable, it may offer more data than is available today if the criteria for its use are changed.

4.6 Predicting the rate and scale of change in the root zone is not possible in advance of a new round of gTLDs

Recommendation 10 - ICANN must develop and document a process for the emergency change related to a temporarily delegated string from the root zone due to collision risk or harms

[83] See "Report a Name Collision," ICANN, accessed 17 January 2024, https://www.icann.org/en/forms/report-name-collision

A new round of gTLDs will require some number of additional delegations to the root zone and workload for IANA. The delegations needed to conduct data collection will only increase that number. As per the RSSAC report to the GNSO Policy Development Process (PDP) Working Group, "the number of TLDs delegated in the root zone should not increase by more than about 5% per month, with the understanding that there may be minor variations from time-to-time."[84] Additionally, the same report described defining a "safe total number of new TLDs" that could be delegated without negative impact to the RSS as a "significant challenge" using only past data. This will likely impact delegation rates, but the extent to which that will be the case is not something analysts can know in advance.

One aspect of an increased rate of change that is not a concern is that of the load on IANA. There have been many changes since the 2012 round, not the least of which is a more efficient set of processes that allows IANA to respond to greater rates of change. IANA's General Manager discussed IANA's capacity with the NCAP DG and reported on the same topic to the GNSO Council at ICANN 78.[85] It is also the case that not all IANA root zone changes needed to support name collision-related data collection will result in new delegations, changes in the size of the root zone, or significant changes in traffic to the root servers. Many of the changes required to implement the data collection methodology discussed in Sec. 3.5 are simply changes to nameserver records, which are lightweight to process and have only small impacts on the size of the zone.

4.7 There is no publicly documented process for emergency changes to the root zone when considering the temporary delegation of strings

Recommendation 10 - ICANN must develop and document a process for the emergency change related to a temporarily delegated string from the root zone due to collision risk or harms

The root zone is critical to the functioning of the DNS, and yet, as far as the NCAP DG is able to determine, ICANN does not have a published, public technical process for emergency changes to the root zone. The Emergency Back-End Registry Operator (EBERO) is designed to protect registrants when a registry operator fails in their contractual obligations, but for individual delegations, no similar process exists. The Root Zone Maintainer Agreement supports a Change Control Process (see Schedule 4) but is limited to coordinating change with the RZM and not the operators of the large public recursive resolvers.[86]

[84] See RSSAC031: Response to the GNSO Policy Development Process (PDP) Working Group on the new Generic Top Level Domains (gTLDs) Subsequent Procedures, https://www.icann.org/en/system/files/files/rssac-031-02feb18-en.pdf

[85] See IANA Update to the GNSO, October 2023, https://static.sched.com/hosted_files/icann78/ef/iana-icann78-gnso-update-202310.pdf

[86] See "Root Zone Maintainer Agreement (RZMA)" - ICANN, https://www.icann.org/iana_imp_docs/129-root-zone-maintainer-service-agreement-v-28sep16.

The NCAP DG identified three potential failure modes that would require an emergency removal of the delegated string:

1. Network Service Provider failure upon delegation - most likely from overwhelming the infrastructure.
2. Major impact to the Internet at-large
3. High-impact to a specific entity(s) that does not create widespread breakage (e.g., one major company is knocked offline or a widely used software package starts having errors).

Should ICANN need to make emergency changes for any reason, there is no publicly documented mechanism to notify the global recursive resolvers or others who may find that information necessary for their operations. No publicly documented process exists to signal to global public resolvers when they need to obtain new copies of root zone data out of their typical schedule.

4.8 The adoption of IPv6 has grown significantly since 2012

Recommendation 8 - ICANN should replace the existing Name Collision Management Framework with the recommended Name Collision Risk Assessment Framework

In 2015, the JAS Report[87] recommended against IPv6 responses during Controlled Interruption because no reliable, universal, and safe equivalent to 127/8 exists in the v6 space, and JAS was concerned that the value (given the exceedingly small number of IPv6-only hosts) did not justify the risk of making something up. The argument at the time was that fewer than 1% of the resolvers sought IPv6-only responses, and only 3.5% of Google users accessed Google services via IPv6. This made sense at the time, but in the intervening years, those numbers have changed significantly.

According to Google's continuous monitoring of IPv6 adoption, just over 40% of Google's users now have IPv6 connectivity.[88] In addition, ICANN announced an IPv6 initiative in 2017 to ensure support for this protocol, at least among ICANN's contracted parties and ICANN org.[89]

4.9 Reserved private-use strings may mitigate the risk of name collisions over the long term but not the short term.

Recommendation 9 - ICANN should create a Collision String List

[87] See Mitigating the Risk of DNS Namespace Collisions: Final Report ("JAS Report"), https://www.icann.org/en/system/files/files/name-collision-mitigation-final-28oct15-en.pdf
[88] See "Statistics," Google IPv6, https://www.google.com/intl/en/ipv6/statistics.html
[89] See ICANN's IPv6 Initiative, https://www.icann.org/resources/pages/ipv6-initiative-2017-02-28-en

As noted in the JAS Report[90], several of the NCAP DG findings, and SAC 113: SSAC Advisory on Private-Use TLDs[91], there is no way to prevent name collisions. As discussed in SAC 113, reserved private-use strings "can help alleviate the uncoordinated ad hoc usage of TLDs for private use." A reserved private-use string is "a domain name label that is explicitly reserved for use as the top-level domain name (TLD) of a privately resolvable namespace that will not collide with the resolution of names delegated from the root zone."

The purpose of a reserved private string is to provide an accepted and agreed-upon target that individuals and organizations can use within their networks for their own purposes.

> Such a reservation should provide a clear path for developers, vendors, service providers, and users to define internally-scoped namespaces for themselves without the requirement for prior coordination, and to do so with the clear understanding that all names in this namespace will never be resolvable in the public Internet, and will not collide with existing or future delegated TLDs in the global DNS.

Establishing private-use space is not a new concept; there is precedent as established by RFC 1918 as it defined private-use, non-routable IP address ranges to help cope with the expected exhaustion of the IPv4 address space.[92] These reserved address spaces are intended for use on local networks only.

While establishing a reserved private-use string may help prevent future name collisions, it is unlikely to have an immediate effect in preventing name collisions. Individuals and organizations must first learn of its existence and establish a practice of using reserved private-use strings as intended.

Additionally, giving preference to strings during the technical review process that are already in private use within internal networks may incentivize private-use strings and result in additional name collision issues. See Section 3.3 on The Issue of Manipulation for further reasoning.

[90] See Mitigating the Risk of DNS Namespace Collisions: Final Report ("JAS Report"), https://www.icann.org/en/system/files/files/name-collision-mitigation-final-28oct15-en.pdf

[91] See SAC113: SSAC Advisory on Private-Use TLDs, https://itp.cdn.icann.org/en/files/security-and-stability-advisory-committee-ssac-reports/sac-113-en.pdf.

[92]See RFC 1918: Address Allocation for Private Internets, https://www.rfc-editor.org/info/rfc1918

5 Recommendations

Given the findings described in this report, the NCAP DG has developed several recommendations for ICANN to work towards in order to offer new gTLD rounds safely and responsibly in a way that is responsive to the issue of name collision. These recommendations should be taken as complementary to the advice found in the New Generic Top Level Domain (gTLD) Subsequent Procedures Policy Development Process Final Report (the “SubPro report”).[93]

5.1 Recommendation 1 - ICANN should treat name collisions as a risk management problem

Finding 4.2.1: Name collisions continue to persist within the DNS

Finding 4.2.2: There are limitations with using currently available data sources for understanding root cause and risk, or designing mitigation and remediation plans

Finding 4.4.1: Critical Diagnostic Measurements are structurally quantitative and benefit from supplemental qualitative information

As discussed in the findings above, there is no single mechanism that will allow ICANN org to identify and mitigate name collisions with a perfect degree of certainty. Nor are there clear quantitative or qualitative measurements that will allow ICANN to determine what type or level of harm (e.g., financial, reputational, or humanitarian) a name collision might be causing. Instead, name collision assessment must be considered a risk management problem.

> Risk management is the process of identifying, assessing and controlling financial, legal, strategic and security risks to an organization’s capital and earnings. These threats, or risks, could stem from a wide variety of sources, including financial uncertainty, legal liabilities, strategic management errors, accidents and natural disasters.
> – IBM, “What is risk management? [94]

Considering name collision assessment as a risk management problem means the ICANN Board must be clear on what level of risk the organization is willing to accept. The acceptable level of risk will inform the risk management process on what data is required to make the necessary assessments. There will be investments required for monitoring, reporting, and detecting name collisions, as well as for responding to and mitigating any name collisions that are discovered.

[93] See Final Report on the new gTLD Subsequent Procedures Policy Development Process (“SubPro Report”), https://gnso.icann.org/sites/default/files/file/field-file-attach/final-report-newgtld-subsequent-procedures-pdp-02feb21-en.pdf

[94] See “What is risk management?” - IBM, https://www.ibm.com/topics/risk-management

All recommendations offered by the NCAP DG depend on the understanding that name collision assessment must be treated as a risk management problem. Each subsequent recommendation works towards determining what data must be collected, how that collection might happen, and how it can be evaluated going forward, as well as how to mitigate any issues discovered.

The validity of an assessment over time is also an assessment that should be considered by the TRT when needed, e.g., when the overall application process for a given string is taking a longer than average length of time.

5.2 Recommendation 2 - ICANN should adopt a consistent definition for name collision

Finding 4.1: The definition of what is a name collision has evolved over time

As noted in Section 1.2, the evolving history around the issue of name collisions has resulted in some variation in the definition of the term "name collision." In order to properly assess the risk and establish the scope of concern, coming to a single, clear definition is critical.

The NCAP DG endorses the following definition:

> Name collision refers to the situation in which a name that is used in one namespace may be used in a different namespace, where users, software, or other functions in that domain may misinterpret it. In the context of top-level domains, the term 'name collision' refers to the situation in which a name that is used in the global Domain Name System (DNS) namespace defined in the IANA root zone may be used in a different namespace (non-IANA), where users, software, or other functions in that domain may misinterpret it.

A complete detailed history of the formal definition of name collisions is provided in the background section of this Report. The above definition has implications regarding the scope of the NCAP study; this is described in detail in Appendix 1 of this report.

5.3 Recommendation 3 - ICANN should continue its education and outreach efforts to the community on the name-collision topic

Finding 4.2.1 Name collisions continue to persist within the DNS

Finding 4.5: Notification to users of name collisions is a critical function and separate from assessment or remediation

The Root Cause Analysis Reports notes that name collision activity has been observed in over half of the TLD strings that have been delegated since August 2014 (when controlled interruption was introduced). This volume of activity was mostly concentrated in a small number of those strings. Nonetheless, the fact that any collision activity was present in so many TLD strings cannot be ignored. While future name collision activity cannot be definitively predicted because of the uniqueness of TLD strings and emergent behavior, general historical observations are the best indicator for predicting future problems. This is an additional reason for ICANN to continue education and outreach.

As noted within Finding 4.5, controlled interruption as a notification method raises awareness of potential name collisions among impacted parties, but this awareness in itself can cause confusion among users who may not understand the risks and consequences of name collisions or the mitigation steps needed to manage name collisions. Hence, currently available methods for notifying affected parties that a name collision has occurred are insufficient for parties to mitigate potential consequences without additional technical assistance and education about name collisions.

ICANN will need to continue to provide education about name collisions for the ICANN community with the goal of raising awareness and preparing the community for the potential of name collisions in the DNS. This recommendation aligns with the outreach campaign ICANN stated it would develop in the New gTLD Collision Occurrence Management Proposal.[95] Additionally, this recommendation reflects the recommendations and implementation guidance available in SubPro's final report.

SubPro Recommendation 13.2 describes the necessity of "an effective communications strategy and plan is needed to support the goals of the [the new gTLD program]."[96] This includes focusing on outreach to applicants, working with the Global Stakeholder Engagement team on disseminating information, and the creation of a single, well-designed website for new gTLD program information. The communications strategy must include information to raise awareness of the possibility of name collisions and the proposed Name Collision Risk Assessment Framework.

[95] See New gTLD Collision Occurrence Management Proposal, https://www.icann.org/en/system/files/files/resolutions-new-gtld-annex-1-07oct13-en.pdf

[96] See Final Report on the new gTLD Subsequent Procedures Policy Development Process ("SubPro Report"), https://gnso.icann.org/sites/default/files/file/field-file-attach/final-report-newgtld-subsequent-procedures-pdp-02feb21-en.pdf

5.4 Recommendation 4 - ICANN should consider the need for mitigation and remediation efforts for high-risk strings

> Finding 4.2.2 There are limitations with using currently available data sources for understanding root cause and risk, or designing mitigation and remediation plans
>
> Finding 4.2.3 .CORP and .HOME demonstrated that high volume is an insufficient measure for analyzing the potential of high-risk impact
>
> Finding 4.4 Quantitative and Qualitative Measurement Considerations

As noted in Finding 4.4, different CDM characteristics will have different implications when assessing risk. A high CDM does not necessarily indicate high risk, nor does a low CDM imply low risk; this is why qualitative review is necessary. Each string must be evaluated independently on a case-by-case basis.

Because of the dynamic nature of the risk assessment, any associated mitigation measures must also be done on a case-by-case basis. Identifying all possible mitigation options is not feasible as every string must be considered based on its own CDMs and appropriate qualitative measures.

To mitigate potential harm related to and also remedy possible name collisions for high-risk strings, the DG has proposed a Name Collision Risk Assessment Framework (See Recommendation 8) that includes the establishment of a Technical Review Team (See Recommendation 7) to review strings for risk level and to appropriately add high-risk strings to the Collision String List (See Recommendation 9) for further review.

5.4.1 Recommendation 4.1 - ICANN should submit .CORP, .HOME, and .MAIL through the Name Collision Risk Assessment Process

> Finding 4.2.3 .CORP and .HOME demonstrated that high volume is an insufficient measure for analyzing the potential of high-risk impact

The ICANN Board has specifically asked for guidance regarding the handling of .CORP, .HOME, and .MAIL. These, as with all strings that have been identified as high risk, should be evaluated according to currently available data using the proposed Name Collision Risk Assessment Process.

5.5 Recommendation 5 - ICANN must support the delegation of strings in order to improve the ability to conduct a name collision risk assessment

Finding 4.2 Name Collision Identification and Quantification

Finding 4.4.2: The quantitative data in CDMs can be improved

The Name Collision Risk Assessment Framework proposed as part of this report is designed to provide insights into name collision risks in incremental actions that will minimize the impact on the community reliant on the NXDOMAIN response currently received from the Root Server System (RSS). Prior to submitting a new TLD application, applicants can examine publicly available systems, such as ITHI and ICANN's DNS Magnitude Page, for name collision activity on the set of strings they are interested in.

In order to gain additional name collision data, a temporary delegation of the applied-for string into the root zone will facilitate the TRT in collecting and measuring additional DNS data at the new authoritative TLD name server. This action effectively simulates an RSS-wide collection of DNS data at the TLD authoritative name server and will also unveil a class of queries that were impaired at the RSS by resolvers implementing privacy-enhancing mechanisms such as QNM. This delegation is part of the workflow proposed in this report and enables the data collection and notification methods described in section 3.5, informed in part by SAC062[97] and the New gTLD Collision Management Proposal[98] regarding mitigation measures that can be taken by using methods similar to "trial delegations.".

In addition to supporting "test delegations" of strings to the root zone, there must be a process for removing strings from the root after test delegation to the root zone when these strings are added to the Collision String List. See Recommendation 9.1: *ICANN should support a mechanism that allows applicants to request a string be removed from the Collision String List*.

[97] See SAC062: SSAC Advisory Concerning the Mitigation of Name Collision Risk, https://itp.cdn.icann.org/en/files/security-and-stability-advisory-committee-ssac-reports/sac-062-en.pdf

[98] See New gTLD Collision Occurrence Management Proposal, https://www.icann.org/en/system/files/files/resolutions-new-gtld-annex-1-07oct13-en.pdf

5.6 Recommendation 6 - ICANN should establish and maintain a longitudinal DNS name collision repository in order to facilitate risk assessments and help identify potential data manipulation

> Finding 4.2.2: There are limitations with using currently available data sources for understanding root cause and risk, or designing mitigation and remediation plans
>
> Finding 4.3.1: There is a risk for CDM data manipulation

As noted in several of the findings shown above (Findings 4.2.2 and 4.3.1), there are a variety of issues with relying solely on the existing datasets for identifying name collisions and their root causes. That said, while existing datasets cannot answer all the questions regarding name collisions, they remain a valuable tool that may help analysts and researchers identify strings at risk for name collision and where CDM data manipulation may be occurring. Longitudinal data may need to be captured to better understand scenarios in which gaming/manipulation of the data might be detectable.

ICANN should continue to invest and extend its measurement systems that provide insights into name collision issues that are readily available to the public prior to any new additional TLD round(s). This may include the extension/expansion of ITHI and further instrumentation of IMRS data. In addition, ICANN org should continue to support such efforts as DITL and facilitate more easily accessible data derivatives from such data collection/analysis efforts. This should also include a history of all name collision assessments, mitigation and remediation plans, and supporting data.

Additional outreach efforts to recursive resolver administrators to establish partnerships for measuring name collisions may be a useful activity to collect data that the IMRS will not see.

5.7 Recommendation 7 - ICANN should establish a dedicated Technical Review Team function

> Finding 4.2.2: There are limitations with using currently available data sources for understanding root cause and risk, or designing mitigation and remediation plans
>
> Finding 4.3.1: There is a risk for CDM data manipulation
>
> Finding 4.3.2: Data manipulation has ramifications beyond the technical aspects of name collision that are influenced by when analysis occurs

The role of ICANN includes coordinating the allocation and assignment of names in the DNS root zone while promoting the security, stability, and resiliency of the DNS. It is critical that ICANN be prepared to restrict name delegation in order to prevent undue harm as a result of high-risk name collisions. It is the responsibility of the Technical Review Team (TRT) function to identify high-risk strings to ensure that their delegation is restricted.

As part of the proposed Name Collision Risk Assessment Framework, the discussion group has recognized the need to have a TRT that will serve four functions: assessing the visibility of name collisions, documenting the results, assessing any mitigation or remediation plans, and implementing an emergency removal of a delegation, if necessary. See Appendix 3 for additional details on the Technical Review Team Development.

Ultimately, the purpose of the TRT is to identify high-risk strings that are problematic. They should be responsible for the reviews of the quantitative and qualitative data available during the gTLD application process. They are also responsible for providing the ICANN Board with advice on gTLD delegation and any need for additional mitigation and remediation. This role should not have operational authority. If the TRT identifies an issue with a delegation, they must contact the IANA function to handle the issue within accepted emergency processes.

To be effective, the TRT must include individuals with significant technical expertise in Internet measurements and the DNS. This function must assess the viability of name collisions, document their findings and recommendations, assess any mitigation and remediation plans, and offer emergency response when necessary. While all members of the TRT should have a basic level of understanding in all of the following areas, the TRT as a whole must have significant technical experience overall.

- Knowledge and understanding of DNS specifications, provisioning, and operation;
- Knowledge and understanding of Internet infrastructure
    - Where it intersects with the DNS;
    - Where it intersects with the usage of the DNS by applications and services;
- Ability to review and understand data collected (e.g., Critical Diagnostic Measurements, or CDMs)
- Ability to understand and assess risk as it relates to the potential for harm

The NCAP DG deliberated extensively on the proposed data collection methods as a small sampling of examples of possible and available methods based upon careful consideration and balance of data privacy risks and potential benefits. The data collection methods proposed for the TRT are a small sampling of known and tested methods. Other methods may be used, but they remain untested and are out of scope within this report. Ultimately, which methods to use should be critically considered during the operationalization of the TRT.

Additionally, time frames for stages of the Name Collision Risk Assessment Framework based on implementation details should be distributed to the public as early as possible.

Given the broad flexibility in implementing the TRT, the NCAP DG does not view it within its remit to provide specific guidance on elements of the operationalization of the Technical Review

Team, including what data to collect, how to assess this data, and how to maintain compliance with data privacy and risk management standards. The intent of not prescribing implementation details is for ICANN org to have broad oversight over these details.

To mitigate delays in the New gTLD Program: Next Round, operationalization of the TRT must be done expeditiously, for which ICANN org would need to provide sufficient resources.

5.8 Recommendation 8 - ICANN should replace the existing Name Collision Management Framework with the recommended Name Collision Risk Assessment Framework

Finding 4.2.2: There are limitations with using currently available data sources for understanding root cause and risk, or designing mitigation and remediation plans

Finding 4.4.2: The quantitative data in CDMs can be improved

Finding 4.5: Notification to users of name collisions is a critical function and separate from assessment or remediation

Finding 4.8: The adoption of IPv6 has grown significantly since 2012

The findings from the various study reports and the input from responses to the Board questions make it clear that a broader set of actions is necessary to acquire the CDMs necessary to inform a name collision assessment. With the collection of data, however, comes the need to analyze said data and offer reasoned advice to the Board. The current Name Collision Management Framework does not adequately address the need to consider name collision as a risk management problem. It therefore must be updated in order to document the need to consider additional quantitative and qualitative data in an evolving Internet.

This risk assessment must be a part of a larger review process for requested strings; ICANN should consider all components of the application process, including the various SubPro requirements, and conduct the name collision risk assessment wherever it considers appropriate. All strings should be subject to a typical technical evaluation process without preferential review treatment for any grouping of strings. The implementation of special procedures for certain types of strings based upon policy adoption is out of scope for this report.

The Name Collision Risk Assessment Framework encourages applicants to review the publicly available data held in datasets such as DITL, the IMRS, and ITHI (see Section 3.2 for more information on what data is available to the public). A review of existing data may provide some insight into the challenges the applicant may face in the formal review process but provides no guarantees or assurances that the string may or may not incur name collisions.

In implementing the Name Collision Risk Assessment Framework, sufficient resources will be needed for expeditious implementation to mitigate delays in the New gTLD Program: Next Round. Additionally, time frames for stages of the Name Collision Risk Assessment Framework based on implementation details should be distributed to the public as early as possible.

When an applicant applies for a new gTLD string, the Technical Review Team (see Recommendation 7) will start the evaluation process with their own review of the publicly available data sets. If, based on the qualitative and quantitative data available, the string is determined to be at a high risk of name collisions that may cause harm, they will recommend to the Board that the string be withdrawn from consideration and added to a Collision String List (see Recommendation 9). If the string is not considered to be at a high risk of name collisions or if the Board requests additional review, the TRT will take additional steps (See Figure 6).

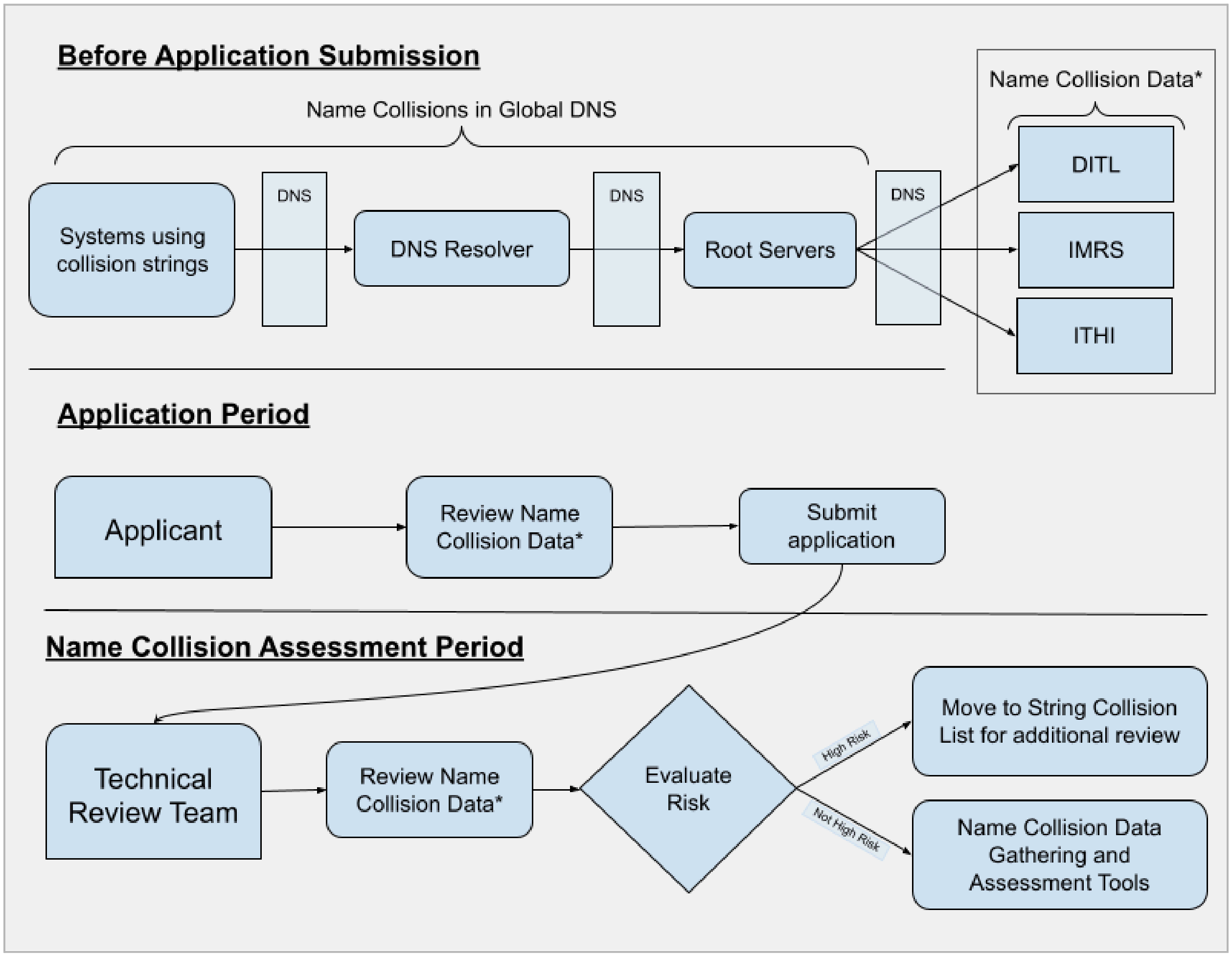

Before Application Submission
Name Collisions in Global DNS
Name Collision Data*
Systems using collision strings
DNS
DNS Resolver
DNS
Root Servers
DNS
DITL
IMRS
ITHI
Application Period
Applicant
Review Name Collision Data*
Submit application
Name Collision Assessment Period
Technical Review Team
Review Name Collision Data*
Evaluate Risk
High Risk
Not High Risk
Move to String Collision List for additional review
Name Collision Data Gathering and Assessment Tools

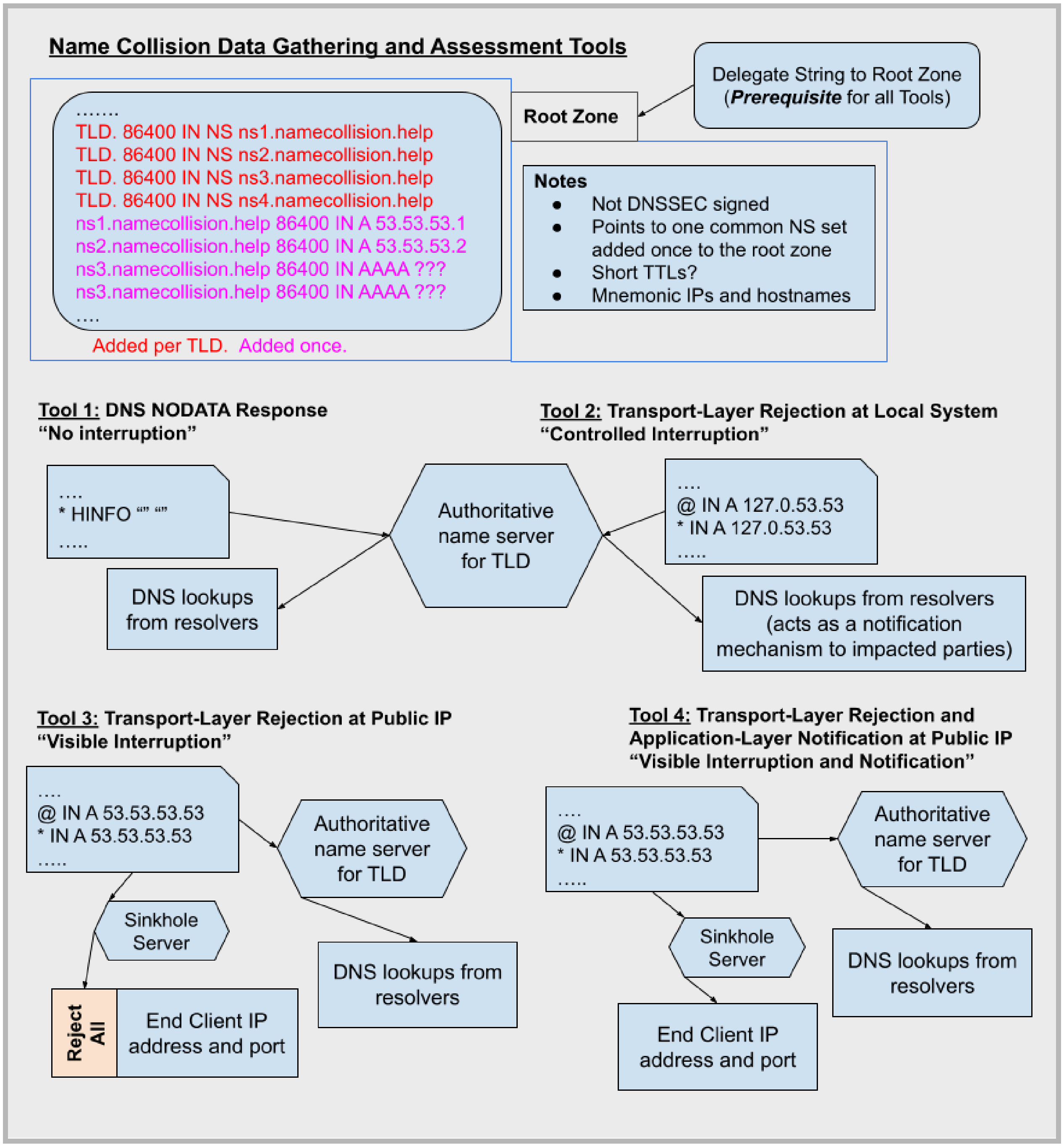
Figure 6: The initial workflow in the proposed Name Collision Risk Assessment Framework

The proposed Name Collision Risk Assessment Framework provides four assessment methods (See Figure 7), described in more detail in Section 3.5, that may be used to collect and assess the data necessary to provide a risk assessment for a given string to the ICANN Board as well as notifying potentially impacted parties.

1. DNS NODATA Response (“No interruption”)
2. Transport-Layer Rejection at Local System (“Controlled Interruption”)
3. Transport-Layer Rejection at Public IP (“Visible Interruption”)
4. Transport-Layer Rejection and Application-Layer Notification at Public IP (“Visible Interruption and Notification”)

Note that DNSSEC should not be used during the trial delegations as it adds unnecessary complexity and does not reflect the behavior of name collisions within the DNS. It would also impair name collision telemetry due to aggressive negative caching.

The NCAP DG deliberated on the proposed data collection methods as a small sample or examples of possible and available methods based upon careful consideration and balance of data privacy risks and potential benefits. These data collection methods are a small sampling of known and tested methods. Other methods may be used, but they remain untested and are out of scope within this report. Ultimately, which methods to use should be critically considered during the operationalization of the TRT.


Name Collision Data Gathering and Assessment Tools
Delegate String to Root Zone (*Prerequisite* for all Tools)
Root Zone
Added per TLD. Added once.
Notes
Tool 1: DNS NODATA Response
"No interruption"
Authoritative name server for TLD
DNS lookups from resolvers
Tool 2: Transport-Layer Rejection at Local System
"Controlled Interruption"
DNS lookups from resolvers (acts as a notification mechanism to impacted parties)
Tool 3: Transport-Layer Rejection at Public IP
"Visible Interruption"
Sinkhole Server
Reject All
End Client IP address and port
Tool 4: Transport-Layer Rejection and Application-Layer Notification at Public IP
"Visible Interruption and Notification"

```
.......
TLD. 86400 IN NS ns1.namecollision.help
TLD. 86400 IN NS ns2.namecollision.help
TLD. 86400 IN NS ns3.namecollision.help
TLD. 86400 IN NS ns4.namecollision.help
ns1.namecollision.help 86400 IN A 53.53.53.1
ns2.namecollision.help 86400 IN A 53.53.53.2
ns3.namecollision.help 86400 IN AAAA ???
ns3.namecollision.help 86400 IN AAAA ???
....
```

- Not DNSSEC signed
- Points to one common NS set added once to the root zone
- Short TTLs?
- Mnemonic IPs and hostnames

```
....
* HINFO "" ""
.....
```

```
....
@ IN A 127.0.53.53
* IN A 127.0.53.53
.....
```

```
....
@ IN A 53.53.53.53
* IN A 53.53.53.53
.....
```

```
....
@ IN A 53.53.53.53
* IN A 53.53.53.53
.....
```

Figure 7: The data collection tools in the proposed Name Collision Risk Assessment Framework

After the data has been collected as per the tools described above, the next step in the Name Collision Risk Assessment Framework is for the TRT to document the data, their analysis, and their recommendation to the ICANN Board. The applicant should also receive a report, which may be adapted according to appropriate privacy policies before distribution.

5.8.1 Recommendation 8.1 - ICANN should not reject a TLD solely based on the volume of name collisions

Finding 4.4.1: Critical Diagnostic Measurements are structurally quantitative and benefit from supplemental qualitative information

Finding 4.2.3: .CORP and .HOME demonstrated that high volume is an insufficient measure for analyzing the potential of high-risk impact

Collecting quantitative data is a critical component of assessing the risk of name collisions, but it must be emphasized that such data is not the only relevant measure. ICANN must be prepared to consider strings that have a high volume of name collisions, as those numbers will not tell the entire story of the risk of harm. During the 2012 round, .CORP and .HOME were examples of strings that required more information than just high volume to understand the impact delegating those strings was likely to have on the DNS.

The problematic nature of measuring harm solely based on CDM values is highlighted by the fact that the Root Cause Analysis Report revealed several strings that:

- Were delegated in the 2012 Round,
- Had higher query volume CDMs than .mail, as noted in the Interisle Report, and
- Received multiple name collision reports via ICANN's reporting form.

Among the 2012 strings with higher CDMs than .mail are the following strings, along with their respective number of ICANN name collision reports:

- Network - 7 ICANN name collision reports
- Ads - 4 ICANN name collision reports
- Prod - 4 ICANN name collision reports
- Dev - 3 ICANN name collision reports
- Office - 1 ICANN name collision report
- Site - 1 ICANN name collision report

5.8.2 Recommendation 8.2 - ICANN should request special attention to strings with high-impact risks during the name collision assessment process

Finding 4.4.1: Critical Diagnostic Measurements are structurally quantitative and benefit from supplemental qualitative information

Finding 4.2.4: It is possible that future name collisions may occur on the scale of .CORP, .HOME, and .MAIL

During the 2012 round, strings that exhibited elevated CDM levels were placed into a category of high risk. Those strings were subsequently investigated to better understand the root cause of the leaking queries and their potential for harm. Unfortunately, the previous name collision and TLD granting workflows did not provide adequate capabilities for applicants and ICANN to abort, terminate, or withdraw applications and place strings into a Collision String List that would prohibit the strings delegation and granting until the string's name collision issues were appropriately mitigated or remediated. In order to address this oversight, the workflow described herein provides a sustainable, repeatable, and deterministic way of assessing name collision risks. As part of that workflow, there are several important opportunities in which strings with high-risk impact warrant additional scrutiny.

Due consideration must be given to those strings that are most at risk from the potential impact as measured by the CDMs throughout the name collision assessment period. In the event of heightened impact risks, the applicant, TRT, and ICANN Board must have an opportunity to reconsider allocation before proceeding with the name collision risk assessment workflow. Decisions made by the TRT or ICANN Board to not proceed should result in the string being placed on a Collision String List.

5.8.3 Recommendation 8.3 - ICANN should update its public-facing name collision reporting process

Finding 4.5.3: The criteria for the use of ICANN's name collision reporting form negatively impacted its use

ICANN currently hosts a web form for individuals to use to report name collisions.[99] This page has significant limits both in terms of what it is intended to collect and its data access policy (i.e., the rules regarding who is allowed to see and use the data collected via that form and for what purposes). Given that the purpose of this form is to help ICANN analyze and understand the

[99] "Report a Name Collision," ICANN, accessed 17 January 2024, https://www.icann.org/en/forms/report-name-collision

source and impact of name collisions, modifying the data policy to allow further research and analysis after the initial submission is necessary.

In addition, the instructions on the form limit its use to individuals who are experiencing "demonstrably severe harm as a consequence of name collision." This limitation should be removed as it may not only deter individuals from reporting suspected name collisions, but it also limits reports collected by ICANN to those that are perceived as posing "a clear and present danger to human life," which is an excessively high ceiling. Changing the requirements for name collision reporting and modifying the text on the web form will allow ICANN to obtain increased reports on name collisions with varying degrees of potential risk or harm. All reports may assist the TRT in evaluating the bigger picture associated with a given name collision.

The TRT must have access to the data from these reports and be free to contact the submitter to request additional information. The form should be explicitly open to any and all name collision reports.

5.9 Recommendation 9 - ICANN should create a Collision String List

Finding 4.2.5: It is impractical to create a do-not-apply list of strings in advance of new requests for delegation

Finding 4.9: Reserved private-use strings may mitigate the risk of name collisions over the long term but not the short term..

While the creation of a do-not-apply list in advance of new requests is impractical for reasons discussed in Finding 4.2.5, there is a need to create a list of strings that the TRT considers high-risk after evaluating them through the proposed Name Collision Risk Assessment Framework in Recommendation 8 (See Figure 8). This list will serve to prevent repeated evaluations until such time as a risk mitigation plan has been proposed and accepted or until other conditions have changed (e.g., a new gTLD round declared or until other technical or policy conditions have changed). The Discussion Group advises that the Board and the Community may need to take steps to consider whether the status of an application listed on the Collision String List should be designated as "Will Not Proceed" or "Not Approved" as further described in SubPro Report 3.4[100].

[100] See Final Report on the new gTLD Subsequent Procedures Policy Development Process ("SubPro Report"), https://gnso.icann.org/sites/default/files/file/field-file-attach/final-report-newgtld-subsequent-procedures-pdp-02feb21-en.pdf

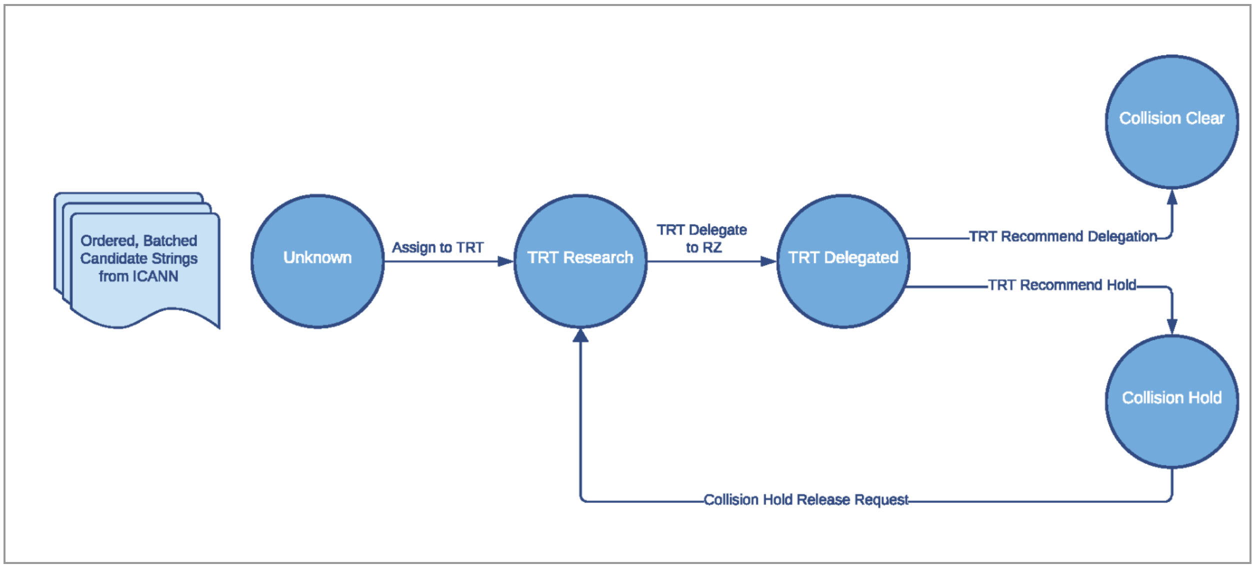


Figure 8: Representation of Technical Review Team workflow for assessing strings

5.9.1 Recommendation 9.1 - ICANN should support a mechanism that allows applicants to request a string be removed from the Collision String List

Finding 4.2.5: It is impractical to create a do-not-apply list of strings in advance of new requests for delegation

In having a Collision String List, there must also be a mechanism to remove a string from that list. As noted in Recommendation 4, however, every string requires a case-by-case evaluation and associated mitigation plan.

The NCAP DG explored several avenues when considering what the process and criteria should be to remove a string from a Collision String List. One option requires the applicant to submit a mitigation plan that is evaluated by the TRT. The TRT then submits a recommendation to the Board as to whether the string may be removed from the list and the applicant allowed to continue or whether the string should continue to be considered high risk and remain in the list.

Another option is to have a process that requires a group similar in governance to the Registry Services Technical Evaluation Panel (RSTEP).[101] It remains an open question as to whether this role might be in place of or in addition to the TRT when it comes to evaluating mitigation plans and recommending a string be removed from the Collision String List.

[101] See "Registry Services Technical Evaluation Panel" - ICANN, https://www.icann.org/resources/pages/technical-evaluation-panel-2012-02-25-en

The NCAP DG looks for guidance from the community as to whether any mitigation plan should be considered on a pass/fail basis versus selecting the best versus determining whether the plan has an acceptable or unacceptable risk level (quantified based on previous evaluations).

5.10 Recommendation 10 - ICANN must develop and document a process for the emergency change related to a temporarily delegated string from the root zone due to collision risk or harms

Finding 4.6: Predicting the rate and scale of change in the root zone is not possible in advance of a new round of gTLDs

Finding 4.7: There is no publicly documented process for emergency changes to the root zone when considering the temporary delegation of strings

The proposed Name Collision Risk Assessment Framework allows for scenarios in which continuing the assessment process results in unacceptable risk to Internet services. For example, a significant surge in the volume and frequency of a name collision might overwhelm the infrastructure of critical network service providers. Another scenario might see a high impact on specific entities (e.g., widely used software packages or large companies knocked offline). If the CDM levels are high enough, there may be an impact on the Internet at large.

In order to be prepared for these and other possibilities of harm due to the delegation of applied-for strings, ICANN must develop and publicize a process for removing a temporarily delegated string from the root zone.

The TRT should not have the operational authority to effect the emergency removal of one of the strings they have delegated as part of the Name Collision Risk Assessment Framework. If the TRT identifies an issue with a delegation, they must contact the IANA function to handle the issue within accepted emergency processes. Additionally, the TRT should be part of the process to assess the request if it comes from an entity other than the TRT itself.

5.11 Recommendation 11 - ICANN should not move ahead with NCAP Study Three

Finding 4.2.2: There are limitations with using currently available data sources for understanding root cause and risk, or designing mitigation and remediation plans

Every new string brings a unique set of CDMs and associated name collision risks. Given the understanding that the currently available data sources and measurement methods are insufficient for understanding designing mitigation and remediation plans, reviewers will need to make decisions on a string-by-string basis based on the best available data and analysis that the TRT has. This makes the development of widely applicable mitigation plans impossible.

As the proposed Study Three is scoped to develop such wide-scale mitigation plans, the NCAP DG recommends that ICANN not move ahead with the third study.

Appendix 1: Revised Definition of Name Collision and Scope of Work

The original RFP for Study One also touched on the possibility of name collisions going beyond the DNS; this was noted as out of scope for the NCAP studies:

> Name collision refers to the situation in which a name that is used in one namespace may be used in a different namespace, where users, software, or other functions in that domain may misinterpret it. In the context of top level domains, the term 'name collision' refers to the situation in which a name that is used in the global Domain Name System (DNS) namespace defined in the root zone as published by the root zone management (RZM) partners ICANN and VeriSign (the RZM namespace) may be used in a different namespace (non-RZM), where users, software, or other functions in that domain may misinterpret it.

However, post-Study One, it was noted by the DG that an item was erroneously included in the "In scope but not intended to be the subject of data studies"[102] as it was in direct conflict with the definition above. Item B.c in which "Registrant Alice uses EXAMPLE.COM and then lets the registration expire. Registrant Bob then registers and delegates EXAMPLE.COM. Traffic intended for Alice's use of EXAMPLE.COM is now received by Bob's use of EXAMPLE.COM". By the definition provided, B.c is out of scope because it must be in a different namespace. A re-registration, by the above definition, is not a different namespace. The resolution process for that name depends on the IANA root zone.

This concern of name collisions is more firmly described in ICANN OCTO's report "Challenges with Alternative Name Systems"[103]:

> "The Domain Name System (DNS) is a component of the system of unique identifiers ICANN helps to coordinate. It is the main naming system for the Internet. It is not the only one. Some naming systems predate the DNS, and others have been recently proposed in the wake of the blockchain approach of decentralized systems.
>
> Proposing a new naming system is one thing. Making sure everybody on the Internet can use it is another. Alternative naming systems face a huge deployment challenge. A number of solutions exist to bridge the DNS to those parallel worlds, but they all come with their own drawbacks.

[102] See Proposed Definition of Name Collisions and Scope of Inquiry for the Name Collisions Analysis Project, published for public comment on 2 July 2019, https://www.icann.org/public-comments/proposed-definition-name-collisions-2019-07-02-en

[103] See Challenges with Alternative Name Systems, https://www.icann.org/en/system/files/files/octo-034-27apr22-en.pdf

> Furthermore, the lack of name space[104] coordination, either between those alternative naming systems and the DNS, or simply among those alternative naming systems, will result in unworkable name collisions. This could lead to completely separate ecosystems, one for each alternative naming system, which would further fragment the Internet.

The NCAP DG therefore endorses the following definition and recommends that the ICANN org adopt a consistent definition for “name collision” (See Recommendation 2):

> Name collision refers to the situation in which a name that is used in one namespace may be used in a different namespace, where users, software, or other functions in that domain may misinterpret it. In the context of top-level domains, the term ‘name collision’ refers to the situation in which a name that is used in the global Domain Name System (DNS) namespace defined in the IANA root zone may be used in a different namespace (non-IANA), where users, software, or other functions in that domain may misinterpret it.

[104] The reference text from which this quote was drawn writes the term “name space” as such.

Appendix 2: Configuration for Notification and Data Generation Methods

No Interruption

```
$TTL 60
$ORIGIN @
@    IN   SOA  ns1.trial-delegation.icann.org. (
               name-collision-admin.icann.org.
                1        ; Serial
                3600          ; Refresh
                3600          ; Retry
               86400          ; Expire
                60 )          ; Negative Cache TTL
     IN   NS   ns1.trial-delegation.icann.org.
     IN   NS   ns2.trial-delegation.icann.org.
*    IN   HINFO "" ""
```

In the above example “@” is replaced with the delegated TLD string. The important parts of the above example are the following:

1. The zone is nearly empty. Aside from the requisite SOA records and NS records, there is only a wildcard HINFO record.
2. The TTL for all records in the zone is 60 seconds, as is the value of the negative cache TTL.

Other aspects of the zone contents, such as the names of servers in the NS records and the MNAME and RNAME fields of the SOA record, can be modified.

The zone contents above do not include DNSSEC records associated with the zone being DNSSEC-signed. Signing the zone with DNSSEC is good practice, but a signed zone makes it subject to aggressive negative caching with NSEC and NSEC3 records. This aggressive caching allows recursive resolvers to infer that a name does not exist without ever issuing a query for that name. This mechanism is efficient, but it results in reduced visibility. If the zone must be signed with DNSSEC, the effects of caching, including aggressive negative caching, can be mitigated, in part, by the 60-second negative cache TTL. Alternatively, a more complex server might be used that supports on-the-fly signing, such as that employed by Cloudflare[105].

[105] See “Economical With The Truth: Making DNSSEC Answers Cheap,” The Cloudflare Blog, https://blog.cloudflare.com/black-lies/

Controlled Interruption

```
$TTL 60
$ORIGIN @
@    IN   SOA  ns1.trial-delegation.icann.org. (
               name-collision-admin.icann.org.
                1        ; Serial
                3600          ; Refresh
                3600          ; Retry
               86400          ; Expire
                60 )          ; Negative Cache TTL
     IN   NS   ns1.trial-delegation.icann.org.
     IN   NS   ns2.trial-delegation.icann.org.
     IN   A    127.0.53.53
     IN   MX   10 your-dns-needs-immediate-attention
     IN   SRV  10 10 0 your-dns-needs-immediate-attention
     IN   TXT  "Your DNS configuration needs immediate attention
               see https://name-collisions.icann.org/"
*    IN   A    127.0.53.53
*    IN   MX   10 your-dns-needs-immediate-attention
*    IN   SRV  10 10 0 your-dns-needs-immediate-attention
*    IN   TXT  "Your DNS configuration needs immediate attention
see https://name-collisions.icann.org/"
```

(Note that the two lines comprising each TXT record should be on the same line for an actual zone file.)

In the above example "@" is replaced with the delegated TLD string. The important parts of the above example are the following:

1. Records of type A, MX, SRV, and TXT exist both at the TLD string itself and as wildcard subdomains of the TLD string.
2. The IP address corresponding to the A records is 127.0.53.53.
3. The record data for the records of the other types contain text referring a user or system administrator to ICANN.
4. The TTL for all records in the zone is 60 seconds, as is the value of the negative cache TTL.

Other aspects of the zone contents, such as the names of servers in the NS records and the MNAME and RNAME fields of the SOA record, can be modified. As noted in section 3.5.2, only A records are used in this configuration; the technique is IPv4-only, as currently proposed. The introduction of a AAAA record for IPv6 support has been proposed but has not been discussed nor tested by the DG.

Visible Interruption / Visible Interruption and Notification

```
$TTL 60
$ORIGIN @
@    IN   SOA  ns1.trial-delegation.icann.org. (
               name-collision-admin.icann.org.
                1        ; Serial
                3600          ; Refresh
                3600          ; Retry
               86400          ; Expire
                60 )          ; Negative Cache TTL
     IN   NS   ns1.trial-delegation.icann.org.
     IN   NS   ns2.trial-delegation.icann.org.
     IN   A    192.0.2.1
     IN   AAAA 2001:db8::1
     IN   MX   10 your-dns-needs-immediate-attention
     IN   SRV  10 10 0 your-dns-needs-immediate-attention
     IN   TXT  "Your DNS configuration needs immediate attention
               see https://name-collisions.icann.org/"
*    IN   A    192.0.2.1
*    IN   AAAA 2001:db8::1
*    IN   MX   10 your-dns-needs-immediate-attention
*    IN   SRV  10 10 0 your-dns-needs-immediate-attention
*    IN   TXT  "Your DNS configuration needs immediate attention
see https://name-collisions.icann.org/"
```

Just as before, “@” is replaced with the delegated TLD string. The important parts of the above example are the following:

1. Records of type A, AAAA, MX, SRV, and TXT exist both at the TLD string itself and as wildcard subdomains of the TLD string.

2. The IP address corresponding to the A records is 192.0.2.1, and IP address corresponding to the AAAA records is 2001:db8::1. Both of these addresses are within the block designated for documentation[106] and are used as placeholders for the actual addresses of the sinkhole server.

3. The record data for the records of the other types contain text referring a user or system administrator to ICANN.

[106] See RFC 5737: IPv4 Address Blocks Reserved for Documentation, https://www.rfc-editor.org/info/rfc5737

4. The TTL for all records in the zone is 60 seconds, as is the value of the negative cache TTL.

Other aspects of the zone contents, such as the names of servers in the NS records and the MNAME and RNAME fields of the SOA record, can be modified.

The contents of the reverse zones for the public IP addresses (192.0.2.1 and 2001:db8::1) used in the Visible Interruption and Visible Interruption and Notification methods include the following:

```
$ORIGIN 2.0.192.in-addr.arpa.
1    IN   PTR
there-is-a-problem-with-your-dns.please-visit.name-collisions.ic
ann.org.
```

```
$ORIGIN 8.b.d.0.1.0.0.2.ip6.arpa.
1.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0   IN   PTR
there-is-a-problem-with-your-dns.please-visit.name-collisions.ic
ann.org.
```

(Note that the two or more lines comprising each PTR record should be on the same line for an actual zone file.)

Finally, the corresponding contents of the zone file for icann.org should include the following:

```
$ORIGIN icann.org
there-is-a-problem-with-your-dns.please-visit.name-collisions IN
A 192.0.2.2
please-visit.name-collisions IN A 192.0.2.2
name-collisions IN A 192.0.2.2
```

(Note that the two or more lines comprising each A record should be on the same line for an actual zone file.)

In this case 192.0.2.2 is a placeholder for an IP address that would host a Web server with more information on name collisions.

Appendix 3: Name Collision Risk Assessment Framework

After considering the variability (i.e., both quantitative and qualitative measures) possible in how to identify name collisions and their potential for harm, the DG considered what the actual workflow might look like in order to evaluate the risks associated with name collisions. Given the goal of a sustainable, repeatable process, the DG iterated on a workflow that ICANN would be able to implement consistently and transparently (See Figure 9). The workflow includes several functions grouped to be executed by a role labeled a Technical Review Team.

The NCAP DG finds it necessary for the ICANN org to maintain broad oversight over implementation details. Therefore, the NCAP DG does provide specific guidance on elements of the operationalization of the Technical Review Team and the Name Collision Risk Assessment Framework, including what data to collect, how to assess this data, and how to maintain compliance with data privacy and risk management standards.

Operationalization of TRT and implementation of Name Collision Risk Assessment Framework should be expeditious, for which ICANN org would need to provide sufficient resources.

Additionally, given the data and privacy risks involved with data collection in general, the ICANN org would need to implement a data privacy and protection policy, along with appropriate risk mitigation measures for legal compliance.

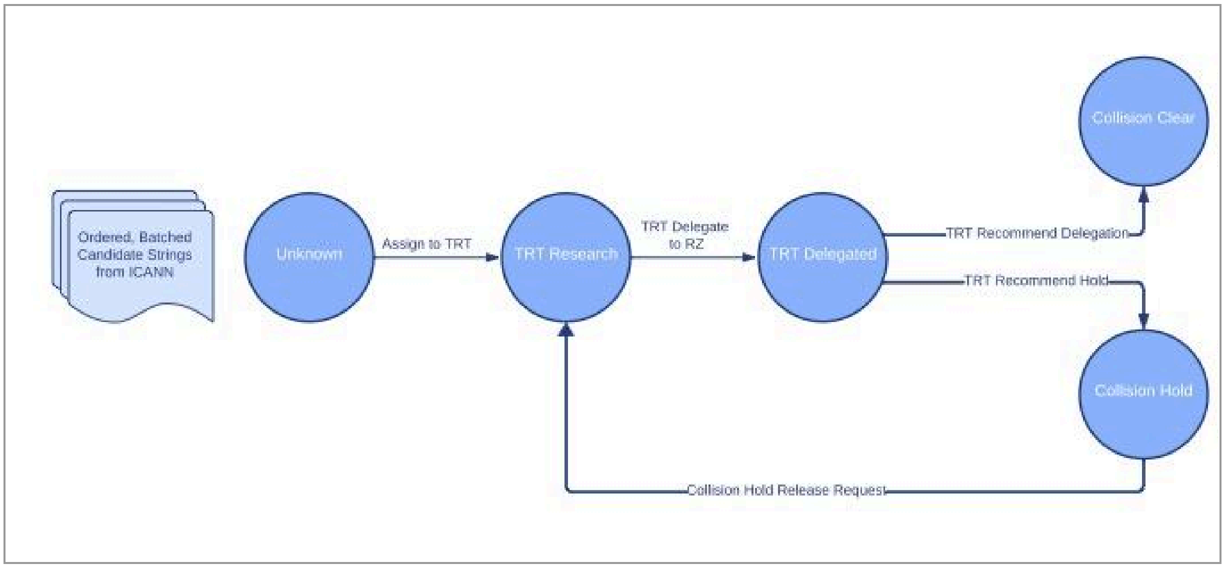


Figure 9: Representation of Technical Review Team workflow for assessing strings

Technical Review Team Development

As part of the proposed name collision workflow, the DG has recognized the need to have a TRT that will serve four functions: assessing the visibility of name collisions, documenting the results, assessing any mitigation or remediation plans, and implementing an emergency removal of a delegation, if necessary. Broadly speaking, members of the TRT are expected to be individuals with significant technical expertise with Internet measurements and the DNS and no conflicts of interest that would impede their neutral evaluation of a delegated string.

While it may be possible for these functions to be handled separately rather than by a single team, for ease of discussion, the DG described all these functions as part of a single TRT's remit. The DG emphasizes that if there is to be a separation of the functions it is essential that all requirements on the composition and execution of the TRT's responsibilities apply to each of the functions.

Assess the visibility of name collisions

The main purpose of the TRT is to identify high risk strings. Their evaluation would happen at various points in time during the application process. At each point, the TRT is expected to document their results as part of making a recommendation to move onto the next assessment activity.

During the initial assessment of Name Collision Data, the TRT would examine the data available prior to the delegation it will request for the next step (e.g., ICANN Managed Root Server (IMRS) logs, ITHI data, DITL data, human-submitted reports, and any other contextual data as may be available) to look for evidence of name collisions. During the evaluation for high-risk strings, the TRT will collect all available CDMs and any other contextual data as may be available, such as unique strings or labels that might help the TRT understand or identify whether a string should be moved to the String Collision List for additional review. The evaluation at this stage is expected to expand over time as the TRT builds a record of previous research. Part of the evaluation would then include comparing the string against a historical baseline to look for known trends.

Strings that are not considered high-risk strings would then be reviewed using Name Collision Data Gathering and Assessment Tools, some of which are included within the report as a small sampling of example available methods. Additionally, the TRT may continue to collect data, including any additional CDMs from protocols other than DNS (e.g., web, email, and others as identified during DNS telemetry gathering) based on the protocols determined during the implementation phase.

Document the results

As noted above, at each point of the evaluation process, the TRT must document their findings to summarize the data seen, measured, and assessed. Any conclusions or recommendations would need to be carefully documented in order to support the goal of transparency.

Part of the documentation effort would include offering reports to the applicant(s) that includes one to two degrees of anonymized, aggregated data. Making this data available allows for an open dialogue with the applicant(s) and should provide insight into any steps needed for developing a mitigation or remediation plan.

At each point, the TRT will be considering what recommendations to make regarding requesting trial delegation, continuing on to deploy selected tools to gather DNS name collision telemetry, and ultimately the final disposition regarding whether or not to recommend awarding the Collision String to the applicant.

The TRT should produce a comprehensive public report on String Collision assessments, actions taken, remediation and risk mitigation plans submitted, along with final determinations, as was done in the 2012 gTLD round.

Assess mitigation and remediation plans

Understanding that mitigation and remediation of name collisions is a case-by-case activity, the TRT is expected to identify when there is a need for such plans. Based on the data they have available from their assessment, they would be in the best position to evaluate how the mitigation and remediation plan offered by the applicant are responsive to the technical issues observed from the CDMs.

Emergency response

When necessary, the TRT would indicate if an emergency response is necessary to revert the delegation at any point in the assessment process (See Recommendation 5). While there is no publicly documented process for the emergency removal of a string test-delegated to the root, the DG determined this is a natural and necessary part of the assessment workflow (See Recommendation 10).

The TRT should understand that its role is to identify high-risk strings that are problematic, i.e., strings that in its technical judgment require a mitigation or remediation (or both) plan(s) prior to allocation.

Evaluation of the Name Collision Risk Assessment Framework

Being able to offer the ICANN Board, or its designee, cogent advice on how to assess the risk of name collisions required the DG to consider what the workflow for such an assessment might look like. The DG focused on the need for a more granular ability to collect data than is possible via the Controlled Interruption process as followed for the 2012 gTLD round. Discussing the workflow, what would be in scope, and what is missing from ICANN's existing policies and procedures took several months (see DG notes from October 2021 through April 2022).

The NCAP DG deliberated extensively, after careful consideration and balance of data privacy risks and potential benefits, on the proposed data collection methods as a small sampling of examples of possible and available known and tested methods. Other methods may be used, but they remain untested and are out of scope within this report. Ultimately, which methods to use should be critically considered during the operationalization of the TRT.

Given the broad variability in operationalizing the TRT and the Name Collision Risk Assessment Framework, the NCAP DG intentionally does not provide specific guidance on elements of the operationalization and implementation of the TRT and Name Collision Risk Assessment Framework, including what data to collect, how to assess this data, and how to maintain compliance with data privacy and risk management standards. The intent of not prescribing implementation details is for ICANN org to have broad oversight over these details.

Sufficient resources would be required for their expeditious implementation, along with guidance on appropriate risk mitigation measures for legal compliance, such as a data privacy and protection policy.

The purpose of the Name Collision Risk Assessment Framework is to identify high risk strings that must include either or both a mitigation and remediation plan intended to reduce the impact of name collisions. The details of that workflow can be found in Recommendation 8.

Each step in the workflow is a linear progression from the previous step; the DG considered it crucial that both the applicant and the TRT be able to place the string into a Collision String List at any step in the process. This option to remove a string from consideration requires the ability for ICANN to do an emergency change to the root zone to remove a delegation (See Recommendation 10).

Process Flow for the Name Collision Risk Assessment Framework

The Name Collision Risk Assessment Framework begins with multiple assessments of a requested string by both the applicant and the Technical Review Team (See Figure 10). For full details, see Recommendation 8.

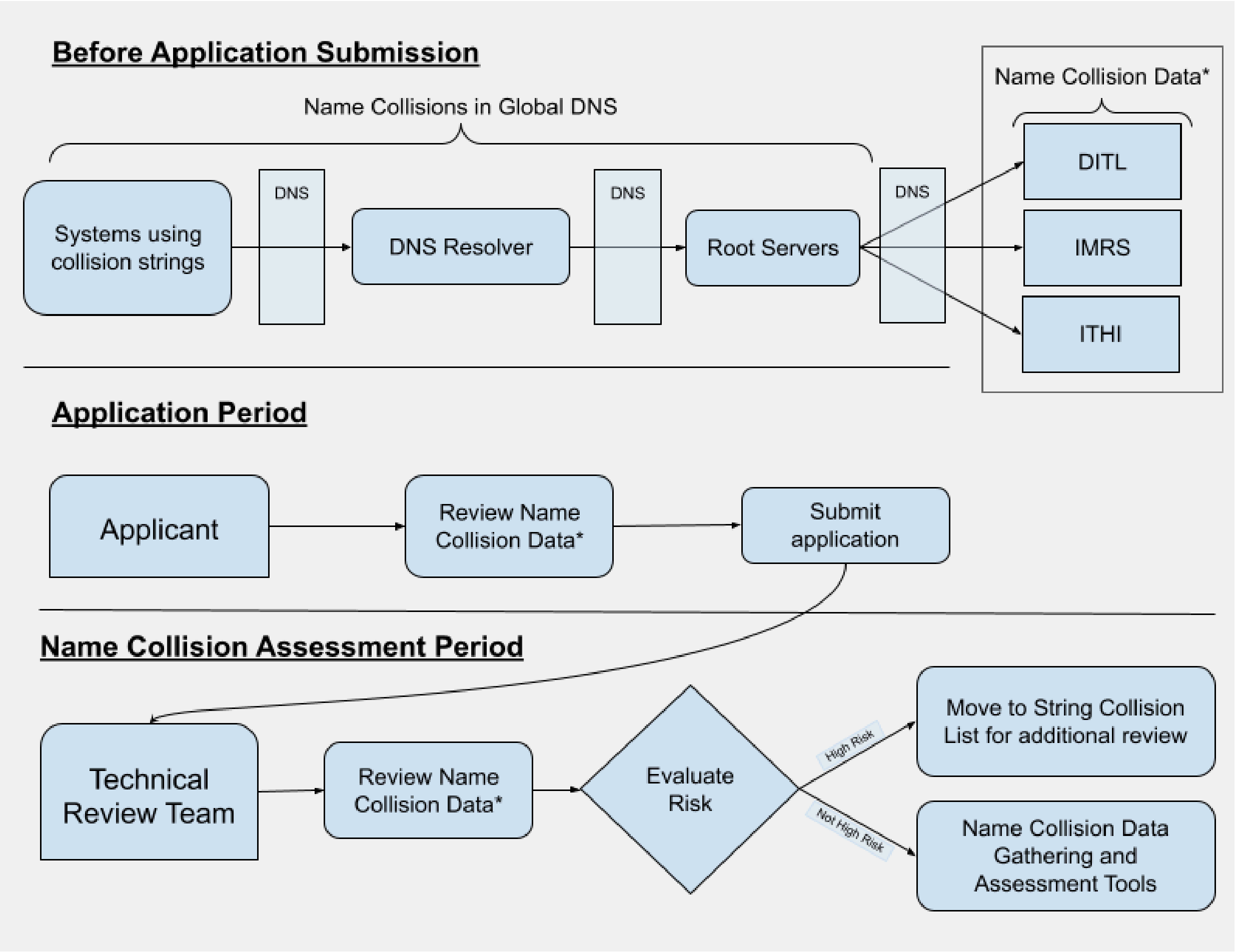

Before Application Submission
Name Collisions in Global DNS
Systems using collision strings
DNS
DNS Resolver
DNS
Root Servers
DNS
Name Collision Data*
DITL
IMRS
ITHI
Application Period
Applicant
Review Name Collision Data*
Submit application
Name Collision Assessment Period
Technical Review Team
Review Name Collision Data*
Evaluate Risk
High Risk
Not High Risk
Move to String Collision List for additional review
Name Collision Data Gathering and Assessment Tools

Figure 10: The initial workflow in the proposed Name Collision Risk Assessment Framework

The proposed Name Collision Risk Assessment Framework provides four assessment methods (See Figure 11). For full details, see Recommendation 8.

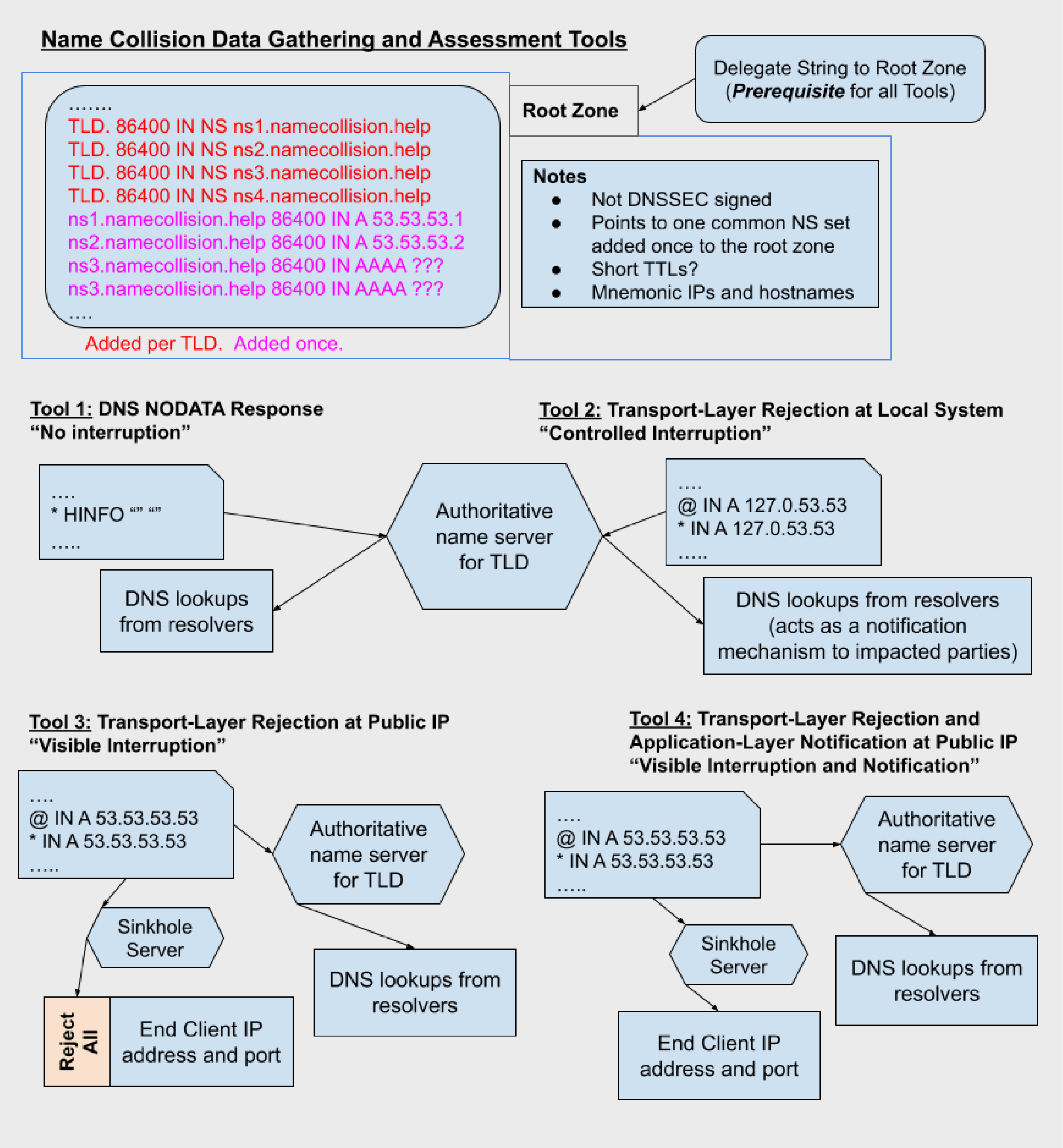

Name Collision Data Gathering and Assessment Tools
Delegate String to Root Zone (Prerequisite for all Tools)
Root Zone
…….
TLD. 86400 IN NS ns1.namecollision.help
TLD. 86400 IN NS ns2.namecollision.help
TLD. 86400 IN NS ns3.namecollision.help
TLD. 86400 IN NS ns4.namecollision.help
ns1.namecollision.help 86400 IN A 53.53.53.1
ns2.namecollision.help 86400 IN A 53.53.53.2
ns3.namecollision.help 86400 IN AAAA ???
ns3.namecollision.help 86400 IN AAAA ???
….
Added per TLD. Added once.
Notes
Not DNSSEC signed
Points to one common NS set added once to the root zone
Short TTLs?
Mnemonic IPs and hostnames
Tool 1: DNS NODATA Response
"No interruption"
….
* HINFO "" ""
…..
DNS lookups from resolvers
Authoritative name server for TLD
Tool 2: Transport-Layer Rejection at Local System
"Controlled Interruption"
….
@ IN A 127.0.53.53
* IN A 127.0.53.53
…..
DNS lookups from resolvers (acts as a notification mechanism to impacted parties)
Tool 3: Transport-Layer Rejection at Public IP
"Visible Interruption"
….
@ IN A 53.53.53.53
* IN A 53.53.53.53
…..
Authoritative name server for TLD
Sinkhole Server
Reject All
End Client IP address and port
DNS lookups from resolvers
Tool 4: Transport-Layer Rejection and Application-Layer Notification at Public IP
"Visible Interruption and Notification"
….
@ IN A 53.53.53.53
* IN A 53.53.53.53
…..
Authoritative name server for TLD
Sinkhole Server
End Client IP address and port
DNS lookups from resolvers

Figure 11: The data collection tools in the proposed Name Collision Risk Assessment Framework

Appendix 4a: Root Cause Analysis - wpad.domain.name

1. Introduction

In 2014, when hundreds of new generic top-level domains (gTLDs) were being introduced into the Domain Name System (DNS), the Internet Corporation for Assigned Names and Numbers (ICANN) introduced a Web-based form by which third parties could report name collisions[107]. Such collisions occur when a domain name is used in a private network environment, but an attempt to resolve that name results in a query to the public DNS. Depending on the nature of the collision and the response to the query in the public DNS, the collision might go unnoticed, it might inhibit legitimate network or application functionality, or it might result in a breach of privacy.

In October 2017, ICANN began receiving reports through its Web form of collisions associated with the domain name `wpad.domain.name.` The reports indicated that HTTP traffic for users in various countries around the world was being proxied through a third party. This man-in-the-middle (MITM) attack violated users' privacy and left them vulnerable to theft of credentials or even identity. The attacks reported resulted from 1) home router software that had a default network configuration, 2) a protocol that made use of that domain to determine where traffic should be directed, and 3) malicious entities that exploited that vulnerability by redirecting traffic to them.

This report was written in direct response to those reports submitted to ICANN. In it we discuss the attack itself and the reports submitted to ICANN. Using artifacts and inferences from historical and recent Internet data, we also create a timeline of events that collectively tell the story of how the network changed over time to create an unsafe environment for vulnerable clients and end users. We also discuss the implications of the circumstances leading to the attack and summarize the key takeaways to be applied to related studies.

2. Background

The Web Proxy Auto Discovery Protocol (WPAD) was proposed in an Internet draft that dates back to 1999[108]. While the draft was never formalized into a Request for Comments (RFC)—the de facto standard for many Internet protocols—it was integrated into nearly every popular Web browser. At the time of writing, Mozilla Firefox and Chrome support WPAD. Additionally, operating systems such as MacOS and Windows offer system-wide proxy settings that include WPAD. Many browsers offer the option of using the system-wide proxy settings in lieu of browser-specific proxy settings. While it is not currently the default setting in many implementations, enabling it is straight-forward and simplifies HTTP proxy configuration.

With WPAD, a browser or operating system discovers an HTTP proxy configuration using one or more methods. One of the most commonly implemented methods involves systematically issuing DNS queries, according to the following pattern. The software retrieves the domain suffix configured on a given system—presumably the domain associated with the organization in

[107] See ICANN, Report a Name Collision Form, https://www.icann.org/en/forms/report-name-collision
[108] See IETF, Web Proxy Auto-Discovery Protocol, https://datatracker.ietf.org/doc/html/draft-ietf-wrec-wpad-01

which it operates. Using that suffix, it forms a domain name by prepending the `wpad` label. For example, the domain name made from the suffix `foo.example.com` would be `wpad.foo.example.com`. An attempt is made to resolve the `wpad` domain name to an IP address. If *not* successful (i.e., because the name doesn't exist or there is no A or AAAA record at the domain name), then the left-most label is removed from the suffix and `wpad` prepended again. For example, a failed attempt at resolving `wpad.foo.example.com` results in an attempt to resolve `wpad.example.com`. This process continues until resolution succeeds. At the point that resolution succeeds, the software issues opens a connection to the IP address to which the name resolved and issues an HTTP request for the URI `/wpad.dat`. The Web server returns a proxy autoconfiguration (PAC) file containing directives, in the form of a script, related to which HTTP proxy server(s) should be used for which clients or Web servers. If the ultimate domain name—and the PAC file retrieved—are managed by a malicious entity, all HTTP requests originated by the software using it can be potentially observed, intercepted, manipulated, redirected, or dropped. This is effectively a man-in-the-middle (MITM) attack. Even HTTPS requests can be interrupted with this configuration. At the very least, the ultimate domain and/or IP address of HTTPS requests made by clients is disclosed to the attacker. In the worst case, the request is intercepted, with the end user being provided a dialog to continue with a connection that is potentially unsafe—which there is a non-zero chance they will click.

We note that WPAD-related vulnerabilities are not new. They have existed as long as the protocol itself[109][110][111][112]. However, the specific situation of `wpad.domain.name` has its own unique story.

3. Vulnerable Configuration Environment

In this section we describe how the combination of a router with an otherwise innocuous default configuration, client devices using WPAD, and an opportunistic domain registration creates a vulnerable network environment for users.

3.1. Home Router Default Domain Name

Home routers often operate a DHCP server. In addition to handing out an IP address, these servers often also distribute a domain suffix. For example, some versions of the D-Link DIR 615 home router provide this suffix to clients as an option that could be configured in the Web console in the "Domain Name" field. However, each router had a *default* suffix, which would be

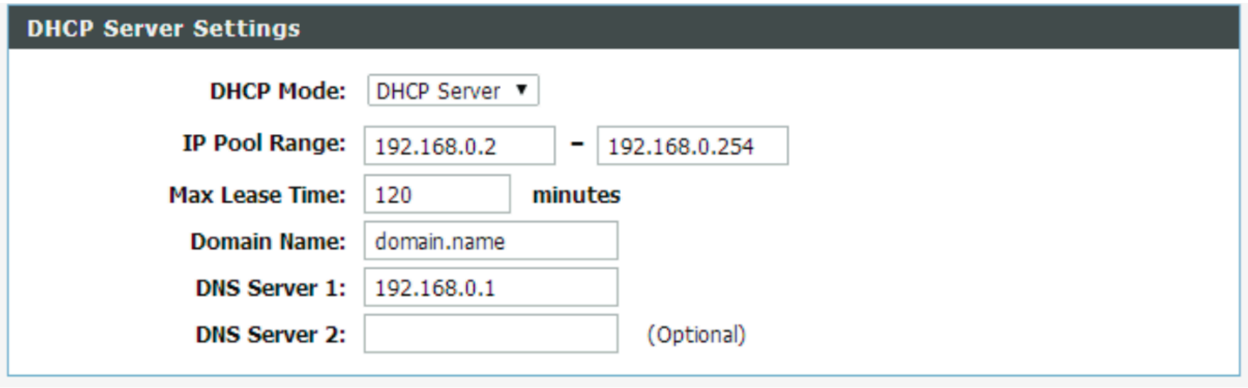
[109] See D. Wessels & M. Fomenkov, "Wow, That's a Lot of Packets," https://www.caida.org/catalog/papers/2003_dnspackets/wessels-pam2003.pdf

[110] See Q. A. Chen, M. Thomas, E. Osterweil, Y. Cao, J. You, & Z. M. Mao, "Client-side Name Collision Vulnerability in the New gTLD Era: A Systematic Study," ACM Conference on Computer and Communications Security (CCS) 2017, https://dl.acm.org/doi/pdf/10.1145/3133956.3134084

[111] See D. Li, C. Liu, X. Cui, & X. Cui, "POSTER: Sniffing and propagating malwares through WPAD deception in LANs," CCS '13: Proceedings of the 2013 ACM SIGSAC conference on Computer & Communications Security (CCS), https://dl.acm.org/doi/pdf/10.1145/2508859.2512520

[112] See Cybersecurity & Infrastructure Security Agency, "WPAD Name Collision Vulnerability," https://us-cert.cisa.gov/ncas/alerts/TA16-144A

distributed to clients unless explicitly changed: `domain.name`. The following image is taken from the manual for the "DIR-615 Revision T3", dated July 11, 2017:



DHCP Server Settings

DHCP Mode: DHCP Server ▼

IP Pool Range: 192.168.0.2 - 192.168.0.254

Max Lease Time: 120 minutes

Domain Name: domain.name

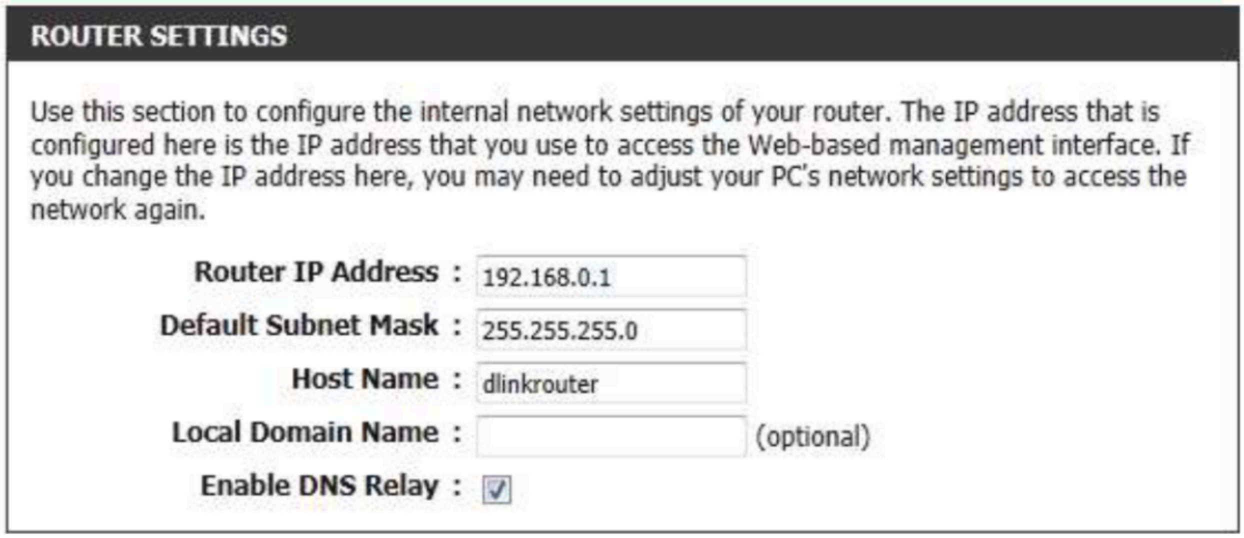
DNS Server 1: 192.168.0.1

DNS Server 2: (Optional)

Source: http://legacyfiles.us.dlink.com/DIR-615/REVT/DIR-615_T3_Manual_v1.10(DI).pdf

The "domain.name" text was likely placed as an (innocuous) example text for the user, providing a description of a value that might appropriately go into the field. Nevertheless, it was and is *used* by clients receiving their network settings from the router. The routers and clients behind them are, of course, not affiliated with the `domain.name` domain name. Thus, using the domain suffix for any protocols, including WPAD, constitutes a *collision*—that is, a domain name being used in a local environment which might coexist with the same name in the public DNS. Unlike name collisions that have been studied at the top-level domain (TLD) level, particularly with the introduction of new generic TLDs (gTLDs)[113], this collision involves a second-level domain, `domain.name`. The `name` TLD has been delegated from the DNS root since 2002.

Not all versions of the D-Link 615 fill in the "Domain Name" (or equivalent) with `domain.name` , as does the previous example. For example, an earlier manual for the D-Link 615 (description: "Initial release"), dated May 20, 2013, shows the following, in which "Local Domain Name" (equivalent to the "Domain Name" field in the 2017 version of the manual):



ROUTER SETTINGS

Use this section to configure the internal network settings of your router. The IP address that is configured here is the IP address that you use to access the Web-based management interface. If you change the IP address here, you may need to adjust your PC's network settings to access the network again.

Router IP Address : 192.168.0.1

Default Subnet Mask : 255.255.255.0

Host Name : dlinkrouter

Local Domain Name : (optional)

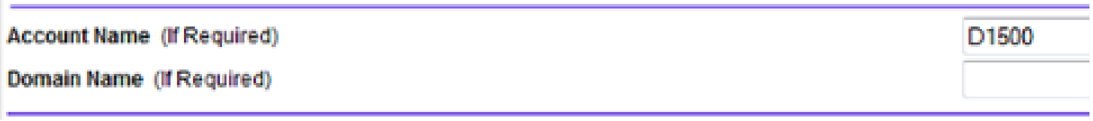
Enable DNS Relay : ☑

Source: https://eu.dlink.com/bg/bg/-/media/consumer_products/dir/dir-615/manual/dir-615_g1_manual_v17_00_eu.pdf

[113] https://www.icann.org/en/system/files/files/name-collision-mitigation-study-06jun14-en.pdf

Similarly, we purchased a new D-Link 615 router for testing. The router version was 5.10 E3. The router, similar to the screenshot of the 2013 version of the manual, did not use `domain.name` in the “Local Domain Name” field, and DHCP responses coming from the router were examined to confirm that in fact no domain suffix was included.

The D-Link 615 is not the only router that is reported to distribute the `domain.name` suffix. The Netgear D1500 Modem Router is also reported to exhibit this behavior (see Section 4.2), although this cannot be confirmed by looking at a May 2018 version of the manual (version 202-11390-02):



Account Name (If Required)	D1500
Domain Name (If Required)	

(Source: https://www.downloads.netgear.com/files/GDC/D500/D500_D1500_UM_EN.pdf)

3.2. Default Domain Name and WPAD: A Dangerous Combination

When a computer system receives its domain name suffix from an affected home router that uses the default configuration of `domain.name`, and software on that system uses WPAD, then the domain name looked up in association with WPAD is very predictable: `wpad.domain.name`. If the name exists, and a PAC file exists at `http://wpad.domain.name/wpad.dat`, then all users behind that router are subject to the HTTP proxy rules found in that file and thus subject to HTTP hijacking.

In summary, the combination of a system that uses WPAD and a home router that hands out a domain for which a third party registers the `wpad` subdomain creates the perfect configuration for a security and privacy vulnerability. Such is the case with `wpad.domain.name`, as will be explained in the next section.

3.3. Delegation and Resolution History of wpad.domain.name

The history of the `wpad.domain.name` domain name, as viewed through various data sources, helps understand the potential client vulnerability over time. We begin our analysis with data retrieved from DNSDB, a historical DNS database generated by Farsight Security from passive DNS feeds[114].

Using DNSDB, we retrieved all DNS records associated with `wpad.domain.name` since 2010. The database includes only records that were observed in responses to recursive-to-authoritative queries where there are DNSDB sensors deployed. While the DNSDB historical records do not show client information, they do include query name (e.g., `wpad.domain.name`), query type (e.g., `NS`, `A`, `MX`, etc.), aggregate query count, time first seen, and time last seen. Additionally, each record includes the “bailiwick” of the server

[114] https://www.farsightsecurity.com/solutions/dnsdb/

responding—an indicator of whether the response came from a child (authoritative) server (e.g., `wpad.domain.name`) or the parent (delegating) server (e.g., `name`).

We divide our assessment into two phases: one in which the delegation appeared to be mostly innocuous, and one in which active man-in-the-middle exploits were observed to take place. The following table contains the history of NS records seen for `wpad.domain.name`, as observed in DNSDB, over both phases.

Dates	**NS Name(s)**	**Parent or Child**	**Number of Responses**
06/2012 to 07/2012	`{a,b,c}.gandi.net`	Both	554
06/2012 to 06/2016	`ns{1,2}.wpad.domain.name`	Parent	9K
07/2012 to 07/2012	`ns{,2,3}.notinuse.notinuse`	Parent	4
07/2012 to 02/2014	`notinuse.notinuse`	Parent	400K
02/2014 to 09/2015	`ns{1,2}.wpad.domain.name`	Child	548
04/2014 to 09/2014	`ns{1,2}.null`	Parent	71K
09/2017 to 10/2017	`ns{,2}.parktons.com`	Both	118K
11/2017 to 07/2021	`ns{1,2}.anycastdns.cz`	Both	5.3M

3.3.1. Phase I - Delegation Only

The domain name `domain.name` is an empty non-terminal; registrations under `name` are always domain names of three labels instead of two. Thus, `wpad.domain.name` is delegated from the `name` TLD. NS records associated with `wpad.domain.name` have been observed in DNSDB as early as June 2012, as shown in the table.

While the initial delegation to `gandi.net` servers was short-lived (about 10 days), a longer-term delegation followed. From June 2012 to June 2016, delegation was observed to `ns1.wpad.domain.name` and `ns2.wpad.domain.name`. During that time, roughly 15 million *referral* responses were observed with these NS names. From February 2014 to September 2015 roughly 500 *authoritative* responses were observed with these NS names. That suggests that either a configuration change occurred in February 2014 causing authoritative responses from `wpad.domain.name` authoritative servers to include NS records in the authority section (i.e., not "minimal responses"[115]) or that client behavior changed such that more NS-type queries increase from 0. With limited other data points, and because the delegation has since changed, it is hard to determine the exact cause.

[115] See https://bind9.readthedocs.io/en/latest/reference.html.

During the same time period in which the NS records for `wpad.domain.name` indicated that it was delegated to `ns1.wpad.domain.name` and `ns2.wpad.domain.name`, *other* NS record sets were also observed in DNS responses. Between July 2012 and February 2014—the same time period that NS records for `ns1.wpad.domain.name` and `ns2.wpad.domain.name` were observed *only* in referral responses–-approximately 440K referral responses were also observed with an NS set composed of the server name `notinuse.notinuse`. Similarly, between April and September 2014, about 71K referral responses were observed containing the NS set having only the names `ns1.null` and `ns2.null`. It is possible that both of these referral responses, composed of NS sets with deliberately unresolvable names, were the result of protective, upstream counter-measures to protect otherwise vulnerable clients from being exploited by the third parties controlling wpad.domain.name. However, it is unclear with the data we have readily available. We investigate this further in Section 4.1.

Several pieces of evidence suggests that this four-year delegation of `wpad.domain.name` was associated with a single registrant. First, the "first observed" and "last observed" dates of the NS records, June 25, 2012 and June 25, 2016, respectively, are consistent with renewal/expiration on an anniversary. Similarly, with the exception of (1) the initial 10-day delegation to `gandi.net` NS names and (2) the curious NS names ending in `notinuse.notinuse`, and `null`, the NS records are consistent throughout the delegation.

The delegation of `wpad.domain.name` from name was apparent from June 2012 to June 2016, as evidenced by the presence of NS records in DNSDB. However, `A` records for `wpad.domain.name` were only observed during the first 10 days of this time period—the 10 days prior to the change in delegation from gandi.net servers to `ns1.wpad.domain.name` and `ns2.wpad.domain.name`. The address to which `wpad.domain.name` resolved during those 10 days, 217.70.184.38, which was within prefixes announced by autonomous system (AS) AS29169, which corresponds to GANDI-AS. This address showed up in responses to a mere 340 queries during those 10 days. Users that observed WPAD-related HTTP requests (i.e., for `http://wpad.domain.name/wpad.dat`) during this time frame reported seeing 404 "not found" responses (see Section 4.2). The reverse DNS entry for 217.70.184.38 is `webredir.vip.gandi.net`, which corresponds to the Gandi parking page[116]. These behaviors and characteristics are consistent with a "domain parking" space. This makes the delegation between June 2012 and June 2016 suspicious but likely innocuous, assuming the resolutions are universally consistent. It is possible that the domain was registered by a registrant that was ignorant of the potential abuse associated with the domain name.

3.3.2. Phase II - Delegation, Resolution, and Interception

For just over a year, from June 2016 to September 2017, no responses were observed containing `wpad.domain.name` records, as observed by DNSDB. Then in September 2017, a new NS set was observed for `wpad.domain.name`, associated with what was apparently a new registration. This claim is supported by the whois information for `wpad.domain.name`,

[116] See also https://gist.github.com/matt-bailey/bbbc181d5234c618e4dfe0642ad80297.

which indicates that `wpad.domain.name` was registered to the current registrant in September 22, 2017, with registration set to expire September 22, 2022. That is, the domain registration was recently renewed.

The initial set of NS records observed in conjunction with this new registration were `ns.parktons.com` and `ns2.parktons.com`. Approximately 108K referral and 9K authoritative responses containing these records were observed over just six days. Immediately following that, beginning in October 2017 and continuing through October 2021, only the following NS records have been observed for `wpad.domain.name`: `ns1.anycastdns.cz` and `ns2.anycastdns.cz`. From October 2017 to July 2021 (date on which historical records were extracted from DNSDB), approximately 3.1M referral and 2.2M authoritative responses were observed. This indicates that there is still currently significant query activity related to `wpad.domain.name`—at least in regions where Farsight Security has placed passive DNS sensors. This is presumably because of the presence of routers running with vulnerable settings (see Section 3.1). As we will discuss in the later section entitled “Firmware Updates", firmware updates have been deployed for at least some home routers affected. This is an indicator not only that these queries are associated with vulnerable routers, but that the routers are vulnerable because they are running outdated firmware.

More significant than the NS records indicating a new delegation of `wpad.domain.name` since 2017 is the fact that A records have been observed for `wpad.domain.name` since that new delegation, whereas they had not been observed previously—other than during the brief 10-day “parking” on Gandi servers in June 2017. A summary of the IP addresses is found in the following table:

Dates	IP Address(es) or /16 IP Prefix(es)	Number of Responses	Autonomous System (AS)	
06/2012 to 07/2012	217.70.184.38	340	AS29169	GANDI-AS
09/2017 to 10/2017	31.192.0.0/16, 159.253.0.0/16	9K	AS43948	GleSYS-AS
11/2017 to 11/2017	51.15.63.145	712	AS12876	ONLINE S.A.S.
11/2017 to 05/2019	91.121.0.0/16, 37.187.0.0/16	4.3M	AS16276	OVH
07/2019 to 01/2020	95.168.185.183	1.6M	AS205544	LEASEWEB UK LIMITED
01/2020 to 04/2020	127.0.0.1	700K	N/A	
04/2020 to 04/2020	94.130.18.141	36K	AS24940	Hetzner Online GmbH
10/2020 to 07/2021	185.38.111.1	2.2M	AS60592	Gransy s.r.o.

During the first six days of the new delegation (i.e., corresponding to NS records in parktons.com), A records mapping wpad.domain.name to IP addresses 31.192.228.197, 159.253.25.197, and 159.253.28.197 were observed. A total of about 9K responses were observed with those IP addresses. All IP addresses were associated—historically, at least—with AS43948, "GleSYS-AS." While it is unclear whether this IP address was used for parking, reports indicate that a simple PAC was being returned when the http://wpad.domain.name/wpad.dat URL was being requested, unlike the similar circumstances when wpad.domain.name was first delegated in 2012 to Gandi servers (see Section 4.2).

From November 2017 to May 2019, wpad.domain.name resolved to between one and three IP addresses, all in AS16276, "OVH." While the set of addresses observed during that 18 months changed twice, the 16-bit prefixes were consistent throughout: 37.187.0.0/16 and 91.121.0.0/16.

For the six-month period between July 2019 and January 2020, wpad.domain.name resolved to the IP address 95.168.185.183, which is associated with AS205544 ("LEASEWEB UK LIMITED"). During this time, approximately 1.5M query DNS responses were observed with that IP address. For the three-month period that followed (January to April 2020), wpad.domain.name resolved to 127.0.0.1. This was possibly a precautionary measure, to interrupt and prevent any malicious activity, but we cannot confirm this with the data. Following that, for a brief six days, wpad.domain.name resolved to 94.130.18.141, an IP address associated with AS24940, "Hetzner Online GmbH". From this latest mapping, about 36K DNS responses were observed.

From October 2020 to the present, wpad.domain.name has resolved to 185.38.111.1, an IP associated with AS60592, "Gransy s.r.o.". From October 2020 to July 2021 (the last DNSDB query associated with this analysis) 2.3M responses were observed associated with this IP address. In the next section we will continue our discussion, noting not only the resolution of wpad.domain.name, but also the content received when HTTP requests were made for http://wpad.domain.name/wpad.dat.

4. Vulnerable Clients - Observations and Reports

From the response counts in the DNSDB entries, we inferred something about the number of queries for wpad.domain.name that have been made during different time periods. However, because those counts are not tied to actual clients, we have no sense for the diversity of the queries. We now use data from additional sources to quantify the pervasiveness of clients potentially vulnerable to man-in-the-middle attack due to vulnerable network configuration.

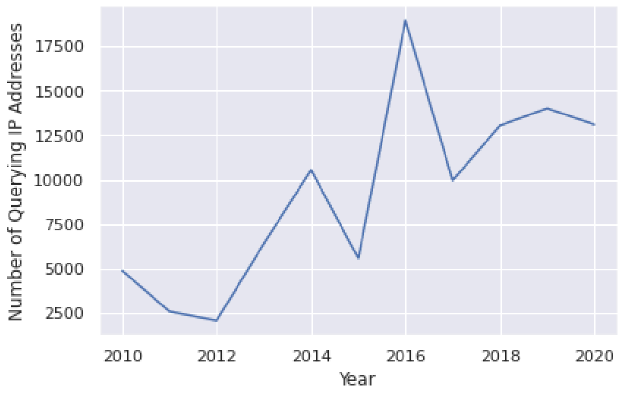
4.1. Queries Observed at the DNS Root servers - DITL

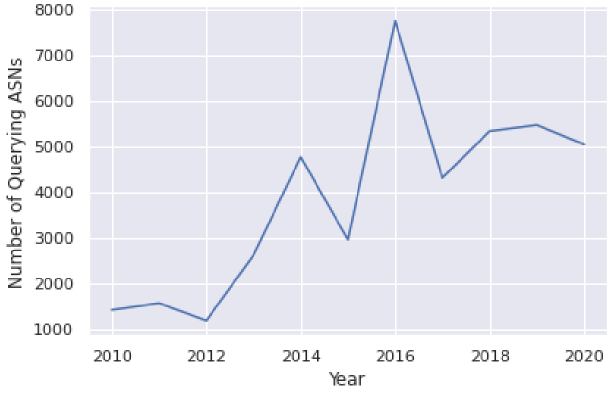
Using the data collected as part of the yearly Day-in-the-Life (DITL) effort, sponsored by the DNS Operations, Analysis, and Research Center (DNS-OARC), and involving most major root server operators, we analyzed DNS queries observed at the root from 2010 through 2020. Each year contains 48 hours worth of captures from all root servers that participated. We extracted the clients and query counts for all queries for `wpad.domain.name` each year for the following root servers: A, C, H, J, K, M. I-root and L-root were excluded every year, even if they participated, because they are known to anonymize client IP addresses—at least in recent years. Some root servers were missing data for some years. Other root letters were excluded because they were not consistent contributors, and their inclusion skewed the query count. H-root was the exception; it was included in our analysis because it participated in every year except 2012. The following table summarizes the data set:

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Avail.	a, b, c, d, e, f, g, h, j, k, m	a, c, d, e, f, h, j, k, m	a, c, e, f, j, k, m	a, c, d, e, f, h, j, k, m	a, c, e, f, h, j, k, m	a, b, c, f, h, j, k, m	a, b, c, e, f, h, j, k, m	a, b, c, d, e, f, h, j, k, m	a, b, c, d, e, f, h, j, k, m	a, c, d, f, h, j, k, m	a, c, d, f, h, j, k, m
No data		b, g	b, d, g, h	b, g	b, d, g	d, e, g	d, g	g	g	b, e, g	b, e, g
Incl.	a, c, h, j, k, m	a, c, h, j, k, m	a, c, j, k, m**	a, c, h, j, k, m	a, c, h, j, k, m	a, c, h, j, k, m	a, c, h, j, k, m	a, c, h, j, k, m	a, c, h, j, k, m	a, c, h, j, k, m	a, c, h, j, k, m
Anon	i, l	i, l	i, l	i, l	i, l	i, l	i, l	i, l	i, l	i, l	i, l

** h is missing

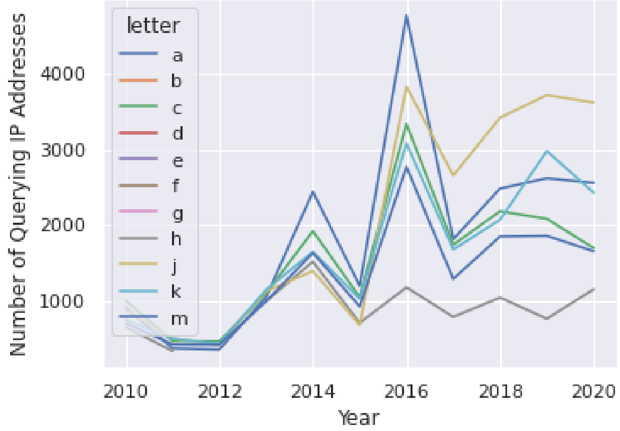
The category “Incl” (i.e., “included”) represents the data used for the rest of our query analysis. A plot of the count of IP addresses and ASNs from which queries originated are shown in the following figure:

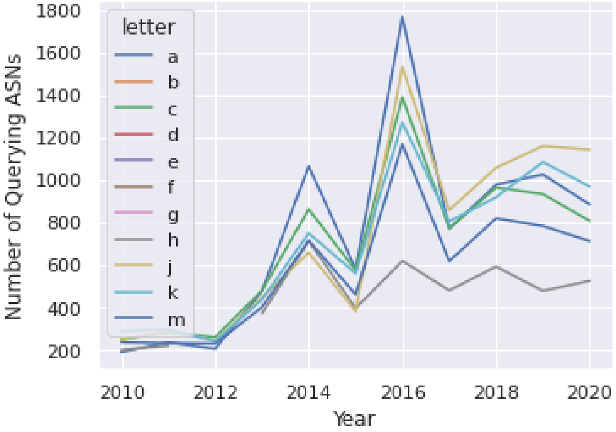




The patterns are remarkably similar, though the raw numbers are different. The years with the lowest client count were 2011 and 2012, in which fewer than 3,000 client IP addresses were observed querying for `wpad.domain.name`, from fewer than 2,000 ASNs. The year with the highest numbers of observed clients was, decidedly, 2016, in which nearly 19K IP addresses queried the root servers for `wpad.domain.name` from almost 9K ASNs, a mean of 2.4 IP addresses per ASN. The DITL collection for 2014 also showed a relatively high number of clients querying for `wpad.domain.name`, both by client IP addresses (about 11K) and ASNs (about 4K).

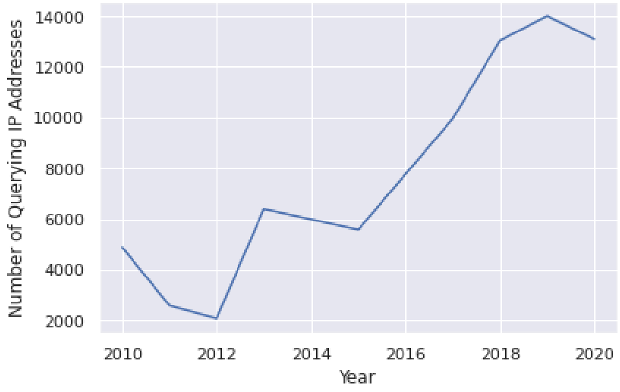
The spikes in 2014 and 2016 are the most obvious features of the graph. We dug further to see if the spikes were a result of bias in the DITL data collection. To test this, we separated the queries for each root letter over the years of the analysis. The resulting graphs follow:

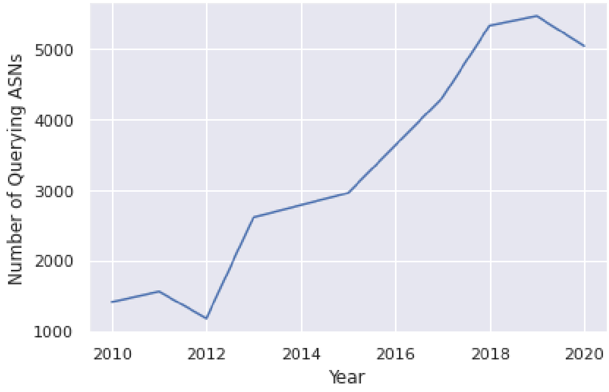




With the exception of M-root in 2016, the relative increase in IP addresses and ASNs issuing queries for `wpad.domain.name` in 2014 and 2016 is observed in all root letters. Thus, whatever the reason for the increases in 2014 and 2016, it does not seem to be due to root letter bias.

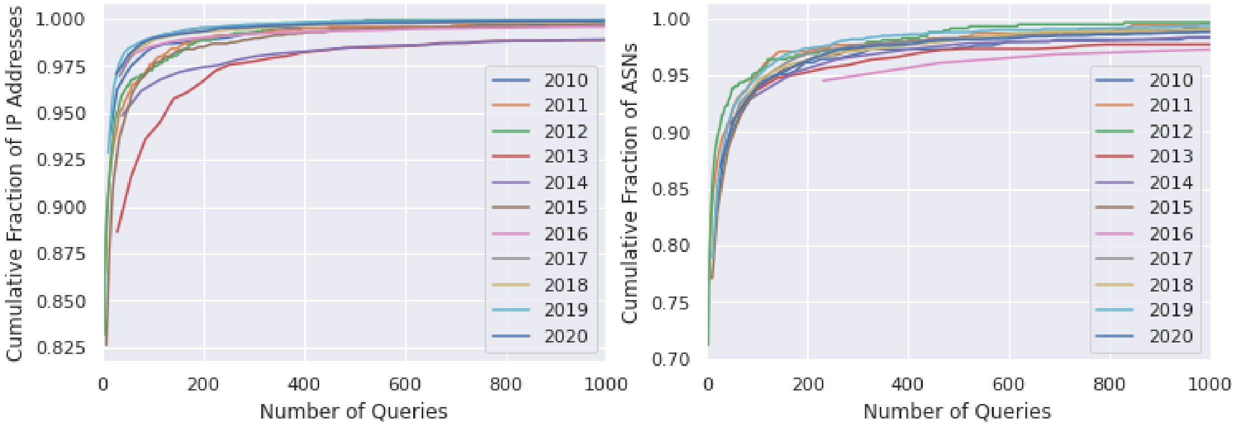
It is difficult to see a clear trend in the plots, when all years are considered. When 2014 and 2016 are removed from the analysis, the trend is a slight upward overall increase since 2012, after which there is a slight decrease, for both client IP addresses and ASNs, as seen from the following plot (2014 and 2016 removed):





Without more data, it is impossible to know how meaningful the spikes in 2014 and 2016 really are and what caused them, and it is hard to tell if the query counts will continue to drop post 2020. All things considered, one thing is for certain. Queries for `wpad.domain.name` are being observed as recent as 2020 from as many as 13K IP addresses and 5K ASNs.

We also plot the distribution of queries for `wpad.domain.name` coming from each IP address and ASN in the following two plots, with the tables containing significant per-IP address and per-ASN statistics following.



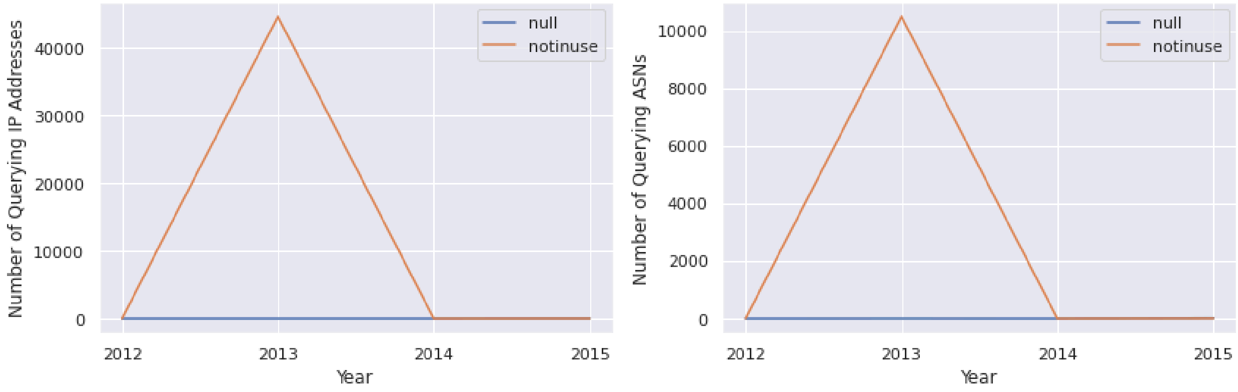
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Med.	1	1	1	1	1	1	1	1	1	1	1
90th Pct.	9	10	10	42	11	18	3	6	5	7	6
Max	14K	4K	3K	28K	38K	6K	31K	33K	52K	10K	26K

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Med.	2	2	1	2	1	1	1	1	1	1	1
90th Pct.	49	33	22	29	23	38	2	7	10	9	3
Max	16K	4K	3K	47K	87K	10K	231K	33K	52K	11K	26K

The median number of queries per IP address for `wpad.domain.name` was 1 for all years. The 90th percentile for number of queries by individual IP addresses was under 20 queries for all years except 2013—that is, 90% of IP addresses issuing queries for `wpad.domain.name` did so fewer than 20 times. Finally, the maximum per-IP address query count over the years analyzed was 52K in 2018.

The median per-ASN query counts have decreased slightly since 2010, from 2 to 1. The 90th percentile per-ASN query counts have likewise decreased from upwards of 49 in 2010 to 3 in 2020.

Up to this point, we have considered queries for `wpad.domain.name` observed at the root servers. We now consider queries for the names `notinuse.notinuse`, `ns1.null`, and `ns2.null`, which corresponded to `wpad.domain.name` `NS` records between July 2012 and February 2014 (`notinuse.notinuse`) and April and September 2014. The number of IP addresses and ASNs from which related queries were received is shown in the following plots:



Only 2013 shows significant query activity and only for `notinuse.notinuse`. This is explained in part by the fact that the 2013 DITL collection was the only one whose date was during the time that `notinuse.notinuse` `NS` records were observed for `wpad.domain.name`, i.e., in DNSDB. *No* DITL collection was carried out during the time that `NS` records containing `ns1.null` and `ns2.null` were observed for `wpad.domain.name`. This explains why the query count for `ns1.null` and `ns2.null` is negligible throughout all years.

Because the client IP addresses typically represent recursive DNS servers, we do not know how many clients—potentially vulnerable—are behind the recursive servers whose behaviors we have analyzed in this section, nor do we know if these queries are actually associated with the D-Link router or more generally with the vulnerability described herein. However, in the next sections we supplement this assessment of potentially vulnerable victims with reports of actual victims of HTTP interception.

4.2. Public Online Reports of wpad.domain.name Interference

We have evidence of name-to-IP-address mappings for `wpad.domain.name` in the DNSDB historical data, and we have evidence of `wpad.domain.name` queries from client IP addresses in the root queries from the DITL data. The mappings tell the story of the *potential* for HTTP interception, and the root server queries are *indicators* of vulnerable clients. However, actual exploitation requires more than DNS queries and mappings; there must be an HTTP response that returns a PAC file directing a system to use a third-party proxy server.

Therefore, the next data we seek is a history of the port 80 responsiveness and HTTP response content corresponding to the URL `http://wpad.domain.name/wpad.dat`. Our questions include the following. Does the system at `wpad.domain.name` even allow TCP connections on port 80? If so, does it respond with a 404 “Not Found” status, a 200 “OK” status, or something else? For a 200 “OK” status, what is the content returned?

The closest thing to an HTTP equivalent for DNSDB is the Internet Archive or “Wayback Machine”[117]. However, the Internet Archive has just a single record for `http://wpad.domain.name/wpad.dat`, dated March 21, 2021, and the content is empty[118]. This behavior is consistent with issuing an HTTP request for `http://wpad.domain.name/wpad.dat` from a single vantage point in the United States, during the time of writing: an HTTP 200 “OK” response with empty response body.

While the Internet Archive has little historical data related to `wpad.domain.name`, and there are no comparable alternatives, there are other data sources. Web-accessible mailing list archives and support forums show reports of interference related to `wpad.domain.name` as early as 2012 and as recently as September 2021. The reports in the archives include historical responses for HTTP requests for `http://wpad.domain.name/wpad.dat`. For example, the following HTTP response was noted on on an engineering and operations group at the Brazilian registry, nic.br on September 27, 2017, and the following day on an ICANN mailing list:

```
      function FindProxyForURL(url, host) {
            return 'PROXY 185.82.212.95:8080; DIRECT';
      }
```

This configuration directs browsers and other HTTP clients using WPAD to connect to 185.82.212.95 port 8080 and issue its HTTP request there as a proxied request, such as:

```
GET http://www.example.com/ HTTP/1.1
Host: www.example.com
```

The contents of the PAC file retrieved at `http://wpad.domain.name/wpad.dat` have changed over time, according to these publicly available but anecdotal reports, which are detailed hereafter. At this point we show the complete history of HTTP response contents, though we modify the whitespace for readability.

The second response observed was reported on a Microsoft mailing list on November 24, 2017. Changes from the previous configuration are shown in red and blue—red for proxy addresses and blue for everything else:

```
      function FindProxyForURL(url, host) {
            if (isPlainHostName(host) ||
```

[117] https://web.archive.org
[118] https://web.archive.org/web/20210325105153/http://wpad.domain.name/wpad.dat

```
                  dnsDomainIs(host, ".windowsupdate.com") ||
                  dnsDomainIs(host, ".microsoft.com") ||
                  dnsDomainIs(host, ".baidu.com") ||
                  dnsDomainIs(host, ".kaspersky.com") ||
                  dnsDomainIs(host, ".live.com") ||
                  isInNet(host, "10.0.0.0", "255.0.0.0") ||
                  isInNet(host, "172.16.0.0", "255.255.224.0") ||
                  isInNet(host, "192.168.0.0", "255.255.0.0") ||
                  isInNet(host, "127.0.0.0", "255.0.0.0"))
            return "DIRECT";
      else
            return 'PROXY 185.93.3.123:8080';
};
```

A third variant, also seen on November 24, 2017, posted on `medium.com`, looked like the previously presented content, but included a different set of proxy IP addresses. Specifically, the line returning the proxy configuration is updated thus:

```
            return 'PROXY 23.111.166.114:8080; PROXY 185.93.3.120:8080';
```

Finally, the following configuration was seen on January 5, 2021 and June 8, 2021, posted to a “Bleeping Computer” forum and a My Broadband forum in South Africa, respectively:

```
function FindProxyForURL(url, host) {
      if (isPlainHostName(host) ||
                  dnsDomainIs(host, ".windowsupdate.com") ||
                  dnsDomainIs(host, ".microsoft.com") ||
                  dnsDomainIs(host, ".baidu.com") ||
                  dnsDomainIs(host, ".kaspersky.com") ||
                  dnsDomainIs(host, ".axaltacs.net") ||
                  dnsDomainIs(host, ".live.com") ||
                  dnsDomainIs(host, ".drivergenius.com") ||
                  isInNet(host, "10.0.0.0", "255.0.0.0") ||
                  isInNet(host, "172.16.0.0", "255.255.224.0") ||
                  isInNet(host, "192.168.0.0", "255.255.0.0") ||
                  isInNet(host, "127.0.0.0", "255.0.0.0"))
            return "DIRECT";
      else
            return 'PROXY 185.38.111.1:8080';
}
```

In addition to the PAC contents returned in the HTTP responses, we summarize the various complaints in the following table:

Date	Country	Router	`wpad.domain.name` IP Address / HTTP Server	HTTP Response	Proxy IP Address / ASN
6/27/2012	Unknown	Trendnet TEW-658BRM	217.70.184.38 / Base HTTP/0.3 Python 2.6	404	N/A

	https://www.wilderssecurity.com/threads/please-help-with-this-outbound-connection-problem.327034/				
9/27/2017	Brazil	D-Link	31.192.228.197, 159.253.25.197, 159.253.28.197	200	185.82.212.95 / AS60592 (Gransy s.r.o.)
	https://eng.registro.br/pipermail/gter/2017-September/071659.html				
9/28/2017	Brazil	D-Link	Unknown	200	185.82.212.95 / AS60592 (Gransy s.r.o.)
	http://mm.icann.org/pipermail/gnso-newgtld-wg-wt4/2017-September/000182.html				
9/28/2017	Unknown	Unknown	37.187.23.23, 37.187.107.197, 91.121.101.78	200	23.111.166.114, 185.93.3.120 / AS29802 (HVC-AS), AS60068 (CDN77)
	https://www.reddit.com/r/networking/comments/732r5n/anybody_else_having_issues_with_wpaddomainname/				
11/24/2017	Unknown	D-Link 890L	Unknown	200	185.93.3.123 / AS60068 (CDN77)
	https://social.technet.microsoft.com/Forums/windowsserver/en-US/e49a45f0-6875-4285-a1d4-5d7de0c63c53/wpad-entry-cannot-browse-websites-using-edge-and-chrome?forum=win10itpronetworking				
11/24/2017	(Maybe) Brazil	D-Link	37.187.23.23, 37.187.107.197, 91.121.101.78	200	23.111.166.114, 185.93.3.120 / AS29802 (HVC-AS), AS60068 (CDN77)
	https://medium.com/@thiago.palmeira/domain-name-wpad-name-collision-exploit-86df7f61d5e5				
1/5/2021	Unknown	Netgear D1500	Unknown	200	185.38.111.1 / AS60592 (Gransy s.r.o.)
	https://www.bleepingcomputer.com/forums/t/740178/was-my-router-compromised-wpad-attack/				
1/9/2021	Italy	ADSL Telecom	Unknown	200	185.38.111.1 /

					AS60592 (Gransy s.r.o.)
	https://www.hwupgrade.it/forum/showthread.php?t=2931491				
1/26/2021	South Africa	Netgear D1500	Unknown	200	185.38.111.1 / AS60592 (Gransy s.r.o.)
	https://mybroadband.co.za/forum/threads/internet-browsing-on-telkom-adsl-not-working-when-check-for-proxy-automatically-is-enabled.1121074/				
6/8/2021	South Africa	D-Link DSL 224 / netis	Unknown	200	Unknown
	https://mybroadband.co.za/forum/threads/pure-dsl-internet-on-laptop-slow-but-fast-on-android.1140307/				
3/24/2021, 9/17/2021	Morocco, others	Netgear D1500	Unknown	Unknown	Unknown
	https://community.kaspersky.com/kaspersky-total-security-14/malicious-object-detected-wpad-dat-wpad-domain-name-trojan-script-agent-dc-merged-16171				

In all cases the IP addresses to which `wpad.domain.name` resolved were consistent with the resolution history reported by DNSDB over the same time frame.

The interference and exploit reported by users and administrators around the world confirmed that there has been a responsive HTTP server at `wpad.domain.name`. Also, there have been some instances of HTTP 404 “Not Found” responses (June 2012) and some instances of HTTP 200 “OK” responses (September 2017 and onward). As for the HTTP 200 “OK” responses, some have returned blank content (such as experienced in our own experimentation and as saved by the Internet Archive), and some have returned content that directs WPAD-enabled systems to use their designated proxy. We further explore this diverse set of HTTP responses in Section 5, specifically looking at how response behaviors differ when HTTP requests are made from different vantage points.

Another observation about the reports is that the geographic regions from which they originate are clustered and do not seem to be representative world-wide. The reports come from Brazil (4), South Africa (2), Morocco (1, containing multiple accounts), and Italy (1). Yet there are no reports from the United States, the United Kingdom, or other countries. We hypothesized that perhaps there were a disproportionate number of vulnerable home routers in the affected countries. That theory is difficult to test. However, in Section 5, we explore another theory which is testable—DNS or HTTP responses that differ depending on the geographic origin of the requests.

4.3. ICANN Name Collision Reports

We now describe the seven reports related to `wpad.domain.name` that came directly to ICANN via the Name Collisions Report Form. We compare them to the reports on public mailing lists and Web forums that we examined in the previous section. We note that there is some bias in comparing them Thus, it is possible, if not likely, that some of the ICANN submissions were made by individuals that were also posting about the issue on public forums.

The dates of the submissions to the ICANN Name Collisions Report Forms are highly correlated with the dates that the postings were made in the public forums. Six of the seven ICANN submissions were made between October 2017 and December 2017. The last submission was made in January 2021, three years later. Of the postings to public forums, five were made between September and November 2017, and three were made in January 2021. Additional online complaints were posted to public forums later in 2021, in June and September. However, the batch of ICANN submissions was retrieved in June 2021, so it is possible that more submissions via the ICANN form have been made since that retrieval.

The country overlap between the ICANN submissions and the public forum submissions is also strong. Four of the ICANN submissions originated from Brazil, one from Italy, and one from the Czech Republic; the origin of the last submission was not provided. In the case of both the ICANN submissions and the postings to public forums, Brazil had the greatest representation, with 3 and 4 reports originating each source, respectively. The ICANN reports from Brazil were dated September through December 2017, and the public forum posts from Brazil were made between September and November 2017. Additionally, Italy was represented in both sets of submissions, with one report from Italy found in each data source.

Finally, there was significant overlap in the devices named in both sources of collision reports. Of the reports submitted to ICANN, three mentioned D-Link, and one of those explicitly mentioned the D-Link DIR 615. Five of the public forum reports included a reference to D-Link devices generally.

5. Present-Day HTTP and Proxy Behaviors

5.1. Behavior and Responses of wpad.domain.name HTTP Server

While historical HTTP response behavior is not available, other than anecdotally, we now report an analysis of current behavior associated with `wpad.domain.name`, as measured from diverse geographic vantage points. Using the Ark platform, made available by the Center for Applied Internet Data Analysis (CAIDA)[119], we issued a DNS lookup for `wpad.domain.name` and an HTTP request for `http://wpad.domain.name/wpad.dat` from 56 vantage points (probes) located in 26 different countries. The DNS and HTTP lookups were all made in September and October 2021.

[119] https://www.caida.org/projects/ark/

Each DNS lookup was performed by issuing a recursive query to the recursive resolver with which each probe was locally configured. The results of the DNS lookup were consistent across all vantage points: in every case, `wpad.domain.name` resolved to the IP address 185.38.111.1. This is the same IP address to which `wpad.domain.name` was observed in the DNSDB history between October 2020 and July 2021 and to which the PAC file at `http://wpad.domain.name/wpad.dat` reportedly directed HTTP clients as an HTTP proxy from January 2021 to present.

While the DNS resolution was consistent from all vantage points, the HTTP response behavior varied. From 50 (89%) of the 56 probes, representing 21 (81%) of the 26 countries, the HTTP response consisted of empty content:

```
HTTP/1.1 200 OK
Date: Fri, 08 Oct 2021 20:06:23 GMT
Content-Length: 0
```

The remaining six probes, from five countries, received the following HTTP response, a slight variant of that most recently reported on public forms (the whitespace has been modified for readability, and changes from the most recently reported contents are highlighted in blue):

```
function FindProxyForURL(url, host) {
      if (isPlainHostName(host) ||
                  dnsDomainIs(host, ".googlevideo.com") ||
                  dnsDomainIs(host, ".youtube.com") ||
                  dnsDomainIs(host, ".windowsupdate.com") ||
                  dnsDomainIs(host, ".microsoft.com") ||
                  dnsDomainIs(host, ".baidu.com") ||
                  dnsDomainIs(host, ".kaspersky.com") ||
                  dnsDomainIs(host, ".axaltacs.net") ||
                  dnsDomainIs(host, ".live.com") ||
                  dnsDomainIs(host, ".drivergenius.com") ||
                  isInNet(host, "10.0.0.0", "255.0.0.0") ||
                  isInNet(host, "172.16.0.0", "255.255.224.0") ||
                  isInNet(host, "192.168.0.0", "255.255.0.0") ||
                  isInNet(host, "127.0.0.0", "255.0.0.0"))
            return "DIRECT";
 else
            return 'PROXY 185.38.111.1:8080';
}
```

These five countries were Japan (2 probes), Mexico, Zambia, South Africa, and Tanzania. The entire list of countries from which HTTP requests were made are shown in the table below:

Country	**PAC content?**	**Country**	**PAC content?**	**Country**	**PAC content?**

Argentina	No	Israel	No	South Africa	**Yes**
Bangladesh	No	Japan	**Yes**	Spain	No
Brazil	No	Madagascar	No	Switzerland	No
Canada	No	Mauritius	No	Tanzania	**Yes**
China	No	Mexico	**Yes**	Ukraine	No
Costa Rica	No	The Netherlands	No	United Kingdom	No
Czech Republic	No	New Zealand	No	United States	No
Germany	No	Paraguay	No	Zambia	**Yes**
Hungary	No	Serbia	No		

The HTTP response behavior is inconsistent over time. The same probes that received HTTP responses with non-empty content days in late September received empty content only days later.

5.2. Behavior of Designated HTTP Proxy Server

We now test the HTTP proxy behavior of the IP address designated by the PROXY string in the PAC file returned by http://wpad.domain.name/wpad.dat. For the 500 top Web sites on the Alexa Top sites, we issued HTTP requests in the following ways:

- An HTTP request directly from our client
- An HTTP request through the proxy
- An HTTPS request directly from our client
- An HTTPS request through the proxy

The objectives with these different requests was to answer the following questions:

- Was the designated proxy server proxying requests generally?
- Was it modifying HTTP requests?
- Was it modifying HTTPS requests?

We make several observations about the results.

The proxy server handles both HTTP and HTTPS requests. HTTP requests are proxied literally—that is, the client issues the HTTP request to the proxy server, the proxy server issues the same request to the Web server, the Web server sends the content to the proxy server as an HTTP response, and the proxy server returns the response to the client. With HTTPS requests the client uses `CONNECT` method with which the proxy server establishes a TCP connection with the Web server over which the client establishes a secure connection using TLS; the HTTP communication happens between client and Web server over an encrypted channel, with the proxy server simply passing along ciphertext.

The proxy server does not tamper with TLS connections. We saw no evidence of MITM wherein a third-party (presumably the proxy server) attempted to impersonate the legitimate Web server when HTTPS was in use. That is, there were no TLS warnings of invalid or even self-signed certificates (except in the few cases where the certificates were actually self-signed).

The proxy server does not modify HTTP responses. Any differences between the content returned from the proxy and that returned by the Web server itself, via direct means, were irrelevant, other than that it was a client with a different source IP address and a different geolocation.

The proxy server modifies HTTP responses under certain conditions. When a Web server exhibits either of the following conditions, the proxy server returns its own HTTP response:

- If the domain name of the Web server does not exist, resulting in an NXDOMAIN rcode. Example: `microsoftonline.com`.
- If the TCP connection to the Web server times out, or is refused (i.e., with a TCP RST). Examples: `163.com` (timeout) and `godaddy.com` (refused).

More particularly, these responses are returned when either of these are the circumstance, as observed by the proxy server itself. At the time of testing, the Web server at the `orange.fr` returned HTTP content to our client (albeit with a 301 HTTP response status), but the proxy returned the proxy's own response content. Subsequent HTTP requests through the proxy returned the Web server's content. We assume that this is because of the proxy server's failure to connect to `orange.fr` at the time of testing. Out of the 500 domains tested, the proxy returned its own content for 29 (5.8%) of the domains, 20 (69%) of which are inaccessible generally, independent of the proxy server, and 9 (31%) of which appear to have been inaccessible to (and and thus the content modified by) the proxy server.

The entire content of this response generated by the proxy server is the following:

```
<html><meta http-equiv="refresh"
content="0;url=http://proxy.domain.name"></html>
```

This has the effect of redirecting the client to the URL `http://proxy.domain.name`. At the time of writing, this URL redirects the client to `https://net.domain.name`. The Web page at `https://net.domain.name` includes just three major links: “Web hosting”, “Create Website”, and “Email Account”. Each link directs the user to a list of ads related to the description of the respective link. Interestingly the site also contains a link to a separate “Privacy Policy” page. This page, last updated in 2014, predates the General Data Protection Regulation (GDPR) both in date and in content. GDPR requires up-front notification to users regarding the use of cookies, with a banner and explicit consent button. The Web site at `https://net.domain.name` does not include the required banner banner, and the privacy page is a generic legal document that includes, among other provisions, the disclaimer that when one visits their Website, they “may track information to administer the site and analyze its usage.” However, there is nothing said about the fact that their original HTTP traffic was intercepted and that contrived content was returned to the client. Nor is there any disclaimer that other HTTP traffic is monitored, even if not modified.

This behavior is the HTTP analog of Site Finder in which a wildcard record was introduced into the `com` and `net` zones[120]. With these wildcard records in place, the `com` and `net` authoritative servers responded to DNS requests for query names with nonexistent second-level domains, such that these domains resolved to IP addresses. These IP addresses listened for and responded to several services, including HTTP and SMTP.

One additional observation is that even in the case where an HTTP response would be contrived by the proxy server for a given domain name (i.e., nonexistent domain, connection timeout, or connection refused), the HTTPS equivalent request (i.e., a `CONNECT` request) would still fail. That is, as long as HTTPS is attempted by the client, no attempt is made by the proxy server to create responses.

5.3. Communication Outreach

In connection with the current research, the CEO of Gransy was contacted to learn more about the delegation, resolution, and HTTP response behavior associated with `wpad.domain.name`. He confirmed that `wpad.domain.name` was registered in 2017 for a so-called “public proxy project”. He indicated that between 2017 and 2021 the PAC content `http://wpad.domain.name/wpad.dat` was mistakenly provided in some countries and that this was corrected after a surge in traffic was noticed or they were notified of a problem. They do not expect it to cause problems in the future. Finally, he indicated that their plan going forward is to only enable the Web server once or twice per year to return empty responses for research purposes.

6. Remediation Efforts

6.1. Public Advisories

The general problem of domain suffixes being used in conjunction with WPAD and the possibility of exploit due to name collision is the subject of a 2016 US-CERT (United States Computer Emergency Readiness Team) / CISA (Cybersecurity and Infrastructure Security Agency) vulnerability announcement[121]. Notably, among the recommendations for those having been exploited in conjunction with the WPAD vulnerability is to report the name collision to ICANN, at the form from which the reports were taken.

[120] https://web.archive.org/web/20041109202247/http://www.verisign.com/static/002702.pdf
[121] https://us-cert.cisa.gov/ncas/alerts/TA16-144A

6.2. Academic Publications

Research supporting the US-CERT announcement was published in the 24th ACM Conference on Computer and Communications Security (CCS) in 2017[122]. This was also discussed in a blog post[123].

6.3. Support Articles

The one support article identified on the Internet, specific to `wpad.domain.name`, is on the Kaspersky support site, posted in May 2021[124]. This was posted in response to the problems on the Kaspersky community support forum, mentioned in Section 4.2. The essence of the support article is to 1) try a different Internet connection, bypassing the router, 2) resetting the router to the default settings, 3) update the firmware to the latest, or 4) stop using the router permanently.

6.4. Firmware Updates

Various updates have been made to the D-Link router firmware over time. However, records of firmware revision history are only found in third-party sites; the D-Link web site lists this as a "legacy product", with the last supported date of January 31, 2018, and no firmware history is shown[125]. The Web site `drivers.softpedia.com` contains the following firmware updates to the D-Link DIR 615, some of which include release notes:

Date	Model	Version	Notes
1/30/2011	DIR-615 Wireless N 300	1.10	- Enhanced Stability. - Updated DDNS UI. - Improved wireless performance.
	https://drivers.softpedia.com/get/FIRMWARE/D-Link/D-Link-DIR-615-Wireless-N-300-Router-Firmware-110.shtml		
5/29/2013	DIR-615 (rev.D)	4.14b02	- **Firmware fixes security vulnerabilities.** - Instructions included.
	https://drivers.softpedia.com/get/FIRMWARE/D-Link/D-Link-DIR-615-revD-Router-Firmware-414b02.shtml		
6/4/2013	DIR-615 (rev.H)	8.04b01	**Fixed publicly disclosed security issues.**

[122] https://dl.acm.org/doi/pdf/10.1145/3133956.3134084

[123] https://nakedsecurity.sophos.com/2016/05/25/when-domain-names-attack-the-wpad-name-collision-vulnerability/

[124] https://community.kaspersky.com/advice-and-solutions-122/what-to-do-if-kaspersky-detects-wpad-17158

[125] https://legacy.us.dlink.com/pages/product.aspx?id=74d82a4d004440a597678377c74080db

	https://drivers.softpedia.com/get/FIRMWARE/D-Link/D-Link-DIR-615-revH-Router-Firmware-804b01.shtml

Two of these firmware releases occurred around the same time in 2013, both indicating that they were security related. We cannot confirm that the security issues listed were a direct reference to the wpad.domain.name vulnerability. Nonetheless they correspond to the time frame during which wpad.domain.name domain was first observed to be delegated with the presence of NS records (June 2012), but before a mapping to IP addresses existed, i.e., with A records (September 2017).

Given that the latest supposed fix for the DIR-615 router was in 2013, and DNS queries for wpad.domain.name have been observed at the root servers through 2020, the combination of one or more other factors might be at play. First of all, it is possible, if not likely, that D-Link routers are running out-of-date firmware, on hardware that is no longer even supported. Second, it is also possible that equipment of other makes or models have similar issues, causing similar symptoms—and that any such devices might have fix dates completely independent of the supposed fix dates for the DIR-615, if fixed at all. Finally, we recognize the fact that not all DNS queries for wpad.domain.name observed at the root represent queries made by stub resolvers to recursive resolvers. In fact, combing through the query noise at the root servers has been a subject of research for many years[126].

6.5. Registration Suspension of wpad.domain.name

In December 2021, the registry operator (Verisign) removed the delegation NS records for wpad.domain.name from the name zone. DNSDB shows that NS records for wpad.domain.name were last observed on December, 9, 2021. Since that time, WHOIS shows the status of wpad.domain.name as “clientHold”, which is “an uncommon status that is usually enacted during legal disputes, non-payment, or when your domain is subject to deletion.”[127] While this status cannot keep vulnerable clients from issuing queries for wpad.domain.name, it can keep them from being exploited, as wpad.domain.name is not delegated and will not resolve to an IP address to which they might otherwise connect.

7. Conclusion

This report details some of the history surrounding wpad.domain.name. As early as 2012, home router implementations, including D-Link’s DIR 615, were configured to distribute the default DNS suffix domain.name to its DHCP clients. When the domain name wpad.domain.name was registered and delegated in 2012, this caused a collision with a name in the public DNS. Prior to 2017, that collision existed without interference to clients

[126] See http://www.sigcomm.org/sites/default/files/ccr/papers/2008/October/1452335-1452341.pdf.

[127] See https://www.icann.org/resources/pages/epp-status-codes-2014-06-16-en#clientHold.

behind vulnerable routers. However, in 2017, a new entity registered `wpad.domain.name`. Since then `wpad.domain.name` has resolved to an IP address, and in many cases that IP address responded to HTTP requests, returning a PAC response with a PROXY string. This PAC directed vulnerable clients to issue all their HTTP requests, with certain exceptions, to the designated proxy server. Thus, HTTP requests from affected clients were at least observed, and, in some cases, intercepted, in a clear violation of privacy and interruption of the end-user Web experience.

The lessons gleaned from this report can serve as a resource for future, related work.
For example, ICANN's Name Collision Analysis Project (NCAP) is being worked on by a team of subject matter experts investigating the past incidence of name collisions, name collision risks associated with the delegation of new gTLDs, and procedures to allow entities to register new gTLDs to be registered with minimal risk. While the current analysis is related to a gTLD (`name`) that is not among those "newly" delegated (i.e., since 2014), it nonetheless has implications applicable to that study and others.

Among the lessons that can be applied generally are the following:

- **Name collisions can occur at any level of the DNS hierarchy.** Incidence of collision at the TLD level has been brought to light because of the relatively new introduction of new gTLDs. However, collisions for names several labels deep might exist, even if the higher-level domain names are already delegated in the public DNS. The current example of this is `wpad.domain.name`; the `name` TLD was delegated in 2002, well before ICANN's new gTLD program.
- **The potential for name collisions might go unnoticed unless and until triggered by an external event.** In the case of `wpad.domain.name`, DNS queries from vulnerable clients were observed years before they were tampered with, opportunistically exploiting their vulnerability. The triggers in this case were the registration of `wpad.domain.name`, and the responses from the `wpad.domain.name` HTTP server directing Web clients to a third-party (i.e., to the clients) HTTP proxy for all subsequent HTTP requests. Had either one of these not happened, clients would be vulnerable to but not be negatively affected by collisions.
- **Firmware in customer premises equipment (CPE) might see delayed updates and security fixes—if it seems them at all.** We assume that the problematic default domain name suffix has been updated in at least some of the firmware, possibly as early as 2013. However, queries for `wpad.domain.name` continue to be observed at the root servers as of the 2020 DITL collection, and online forum posts indicate collisions with `wpad.domain.name` as recently as September 2021. This highlights a problem with CPE devices.
- **Users affected by name collisions might not know what is going on or where to report the problem.** Some users affected by the `wpad.domain.name` issues posted to online support forums, and others reported the issues to ICANN via ICANN's online submission tool, which they likely found—in this instance with a Web search. Those end users that posted to online forums received feedback from other users or support

representatives and, in many cases, were able to resolve their issues. However, in either case—and more generally, there was no definitive way for users to know how their system was affected, who was responsible, and who to contact to get their system back up and running and/or shut down any nefarious activity.

- **Name collisions might not affect users universally.** In this case, HTTP requests were treated (i.e., responded to) differently depending on the country or region associated with the IP address from which the HTTP request originated. Whether the reason was to balance the load, target users geographically, or confuse investigators, or whether it was simply accidental, we may never fully know. However, the resulting behaviors made its investigation more challenging.

We hope that this report serves the purpose for which it was written—both to provide a better understanding of the vulnerabilities and exploits of affected clients and to provide thoughtful discussion for similar, future circumstances.

Appendix 4b: Root Cause Analysis - New gTLD Collisions

1. 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.”[128] 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 primarily takes into consideration 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[129];
- 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.

2. Background

This section provides technical background related to our study.

2.1. DNS Suffix Configuration

The network configuration for most operating systems includes the option for a DNS “suffix” (e.g., `example.com`) to be specified for various purposes. The system’s stub resolver library,

[128] https://community.icann.org/display/NCAP/
[129] https://www.icann.org/en/forms/report-name-collision; Appendix A.

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 2.4) or potential routers for IPv6-over-IPv4 tunneling (see Section 2.5).

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.

2.2. Controlled Interruption

Some systems 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 w*ithi*n 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* application behavior.

Controlled interruption involves inserting *wildcard* records in the otherwise empty zone file associated with a previously undelegated TLD. The wildcard `A` (IPv4 address) record in the zone file maps to a non-routable address: 127.0.53.53. Thus, *any* `A`-type 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. Note that there is no IPv6 equivalent for queries of type `AAAA` (IPv6 address).

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.

2.3. Chrome Browser NXDOMAIN Probing

On startup, the Google Chrome Web browser historically issued three queries, appending the system DNS suffix (see Section 2.1) 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.

2.4. 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 appends the DNS suffix with which a system is configured (see Section 2.1) 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.

2.5. 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 2.1) 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`. During the controlled interruption period for a given TLD, all ISATAP-related queries under the TLD result in positive DNS responses.

3. Data Sets

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

3.1. 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[130]. Each report included, among other information, the date of the report, the TLD in question, a brief description of the problem, and contact information of the submitter. The entire form is included in Appendix A. We use the data from these reports to better understand user and organization experience associated with the delegation of new TLDs in Section 4.

[130] https://www.icann.org/en/forms/report-name-collision

3.2. DNSDB

DNSDB, operated by Farsight Security (part of DomainTools), 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, the database is limited to network locations where sensors are deployed. Also, if an observed query for a given name results in a negative response (i.e., no mapping), DNSDB will have no entry for that name.

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. We describe each in the following sections.

3.2.1. Controlled Interruption Queries

We used ICANN's published list of delegated strings[131] 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—and observed by passive DNS sensors—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.

[131] https://newgtlds.icann.org/en/program-status/delegated-strings

3.2.2. 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 the subset of namespace under each TLD that was associated with name collision activity. This namespace is identified in Section 6 and refined in Section 8, ultimately resulting in 2,266 subdomains associated with 166 of the new TLDs. 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.

3.3. 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 the query name and querying IP address for all queries associated with the 2,266 subdomains that we identify in Section 8 for DITL collections 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 (not all root letters are represented in all years). We refer to this data set as DITL-2014-2021. This is further described in Section 8.

3.4. Web Search Results

We performed two Web searches using Google’s search engine. We searched for the following search terms between September 28 and October 4, 2022: “controlled interruption” and “127.0.0.53”. Each search term included the quotation marks. In each case, we looked at the first six pages of search results. For the search term “controlled interruption,” every result was either completely unrelated to the controlled interruption implemented by ICANN, or it involved documentation or announcements involving controlled interruption. For the search term “127.0.53.53,” we observed 17 results that described unique cases in which controlled interruption was experienced in such a way that the normal flow appeared to be disrupted. If the page in the result appeared to convey a matter of mere curiosity about behavior, rather than disruption, then we excluded the results. The full set of results are found in Appendix B. An analysis of the results is found in Section 5.

4. Name Collisions Report Analysis

We now analyze the reports submitted to ICANN via the Web form (see Section 3.1). We note that this data set has inherent bias in three ways. *First*, the submission of the report itself *implies* that a user or organization was impacted in some way by name collisions, so we cannot suggest that the reports herein are the *only* experience that was had; it is possible that some users of private DNS namespace were not impacted and that their story is not captured. *Second*, the submission implies that they found the online form. This means that the question of the *effectiveness* of the use of the controlled interruption IP address (127.0.53.53) in helping the user or administrator 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*. *Finally*, ICANN's Web form invites users to submit a report only if they are "suffering demonstrably severe harm as a consequence of name collision." Thus, some users impacted by name collisions—but not in such an extreme way as described by the form instructions—might have been dissuaded from submitting a report at all. Later in the paper (see Section 9) we describe a survey sent to 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.

4.1. 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.

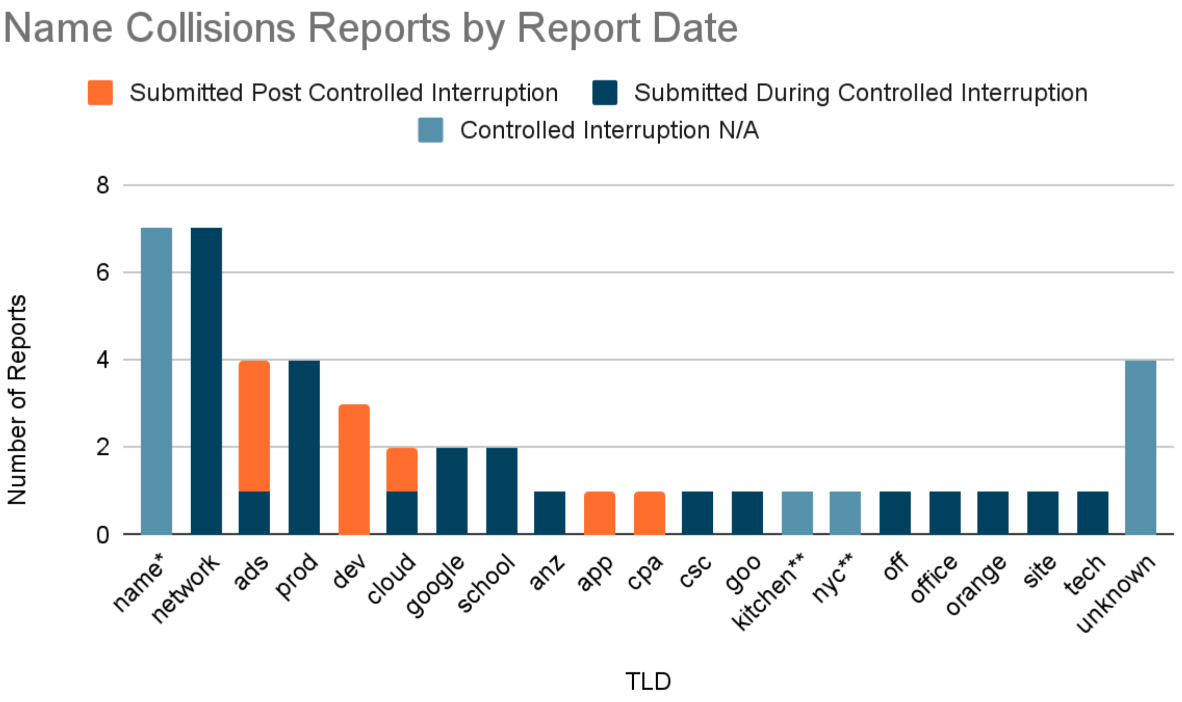
Category	**Count**	**Subcat. %**	**Total %**
Total reports	**47**	**100%**	**100%**
do not include TLD	4	8.5%	8.5%
include TLD	**43**	**91%**	**91%**
delegated prior to new TLD program*	7	16%	15%
delegated as part of new TLD program	**36**	**84%**	**77%**
prior to controlled interruption (pre-Aug 2014)**	2	6%	4.3%
with controlled interruption (Aug 2014 or later)	**34**	**94%**	**72%**
report date is during controlled interruption	25	74%	53%
report date is post controlled interruption	9	26%	19%
reported by organization	24	71%	51%
reported by individual	9	26%	19%
reported origin unknown	1	3%	2.1%
Total TLDs reported	**20**	**100%**	**100%**
delegated prior to new TLD program*	1	5%	5%
delegated as part of new TLD program	**19**	**95%**	**95%**

prior to controlled interruption (pre-Aug 2014)**	2	11%	10%
with controlled interruption (Aug 2014 or later)	17	89%	85%

Each percentage in the “Subcategory %” column is taken from the “Count” in the “Parent” category or subcategory (i.e., the bolded count most immediately above). The percentages in the “Total %” column are taken from the “Count” in the “Total TLDs” or “Total Reports” category.

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. Other TLDs mentioned are `name`, `nyc`, and `kitchen`. The `name` TLD, delegated before the new TLD program (*), was associated with 7 reports. All 7 reports were associated with the delegation of `wpad.domain.name`, which allowed the HTTP traffic of affected parties to be monitored and intercepted by third parties. This is discussed in a separate report (ref report). The `nyc` and `kitchen` TLDs were delegated as part of the new TLD program prior to controlled interruption**.

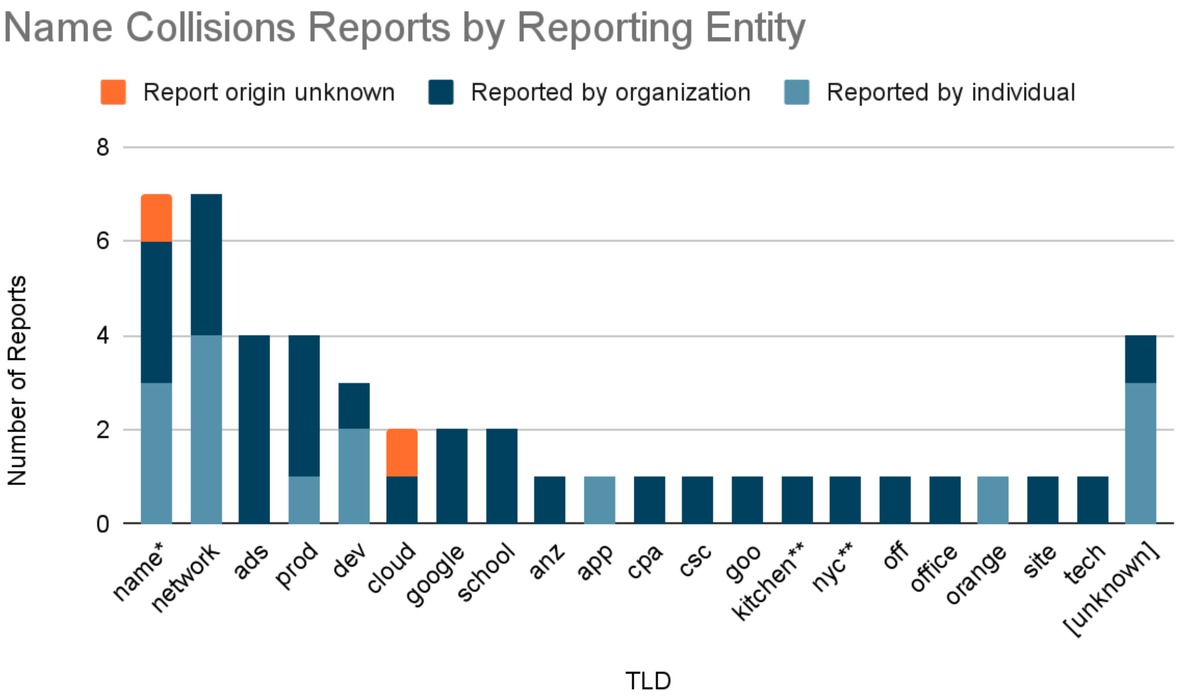
The following plot shows the distribution of reports by TLD, including those that were not delegated as part of the new TLD program (*) and those that were delegated prior to controlled interruption (**). For the 17 reported TLDs that were delegated after controlled interruption was introduced, each 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. With the exception of `cloud`, all TLDs for which reports were received after the controlled interruption period were observed using the controlled interruption IP address (127.0.53.53) beyond the designated time: an additional 693 days for `ads`, 1,042 days for `dev`, 593 days for `app`, and at least 644 days for `cpa`. (Domain names within the `cpa` TLD were still resolving at the time we retrieve the historical data.) See Section 7 for more. In every one of these cases, the report date was prior to the date that the controlled interruption IP address was last observed for the TLD in question.

4.2. 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 (**).



Two standouts are `ads` and `school`, for which reports were made exclusively by organizations. The `ads` TLD (as well as `local` and `intern`) is reportedly (as indicated in one report, but not independently verified) used in books and training resources 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.

4.3. 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.

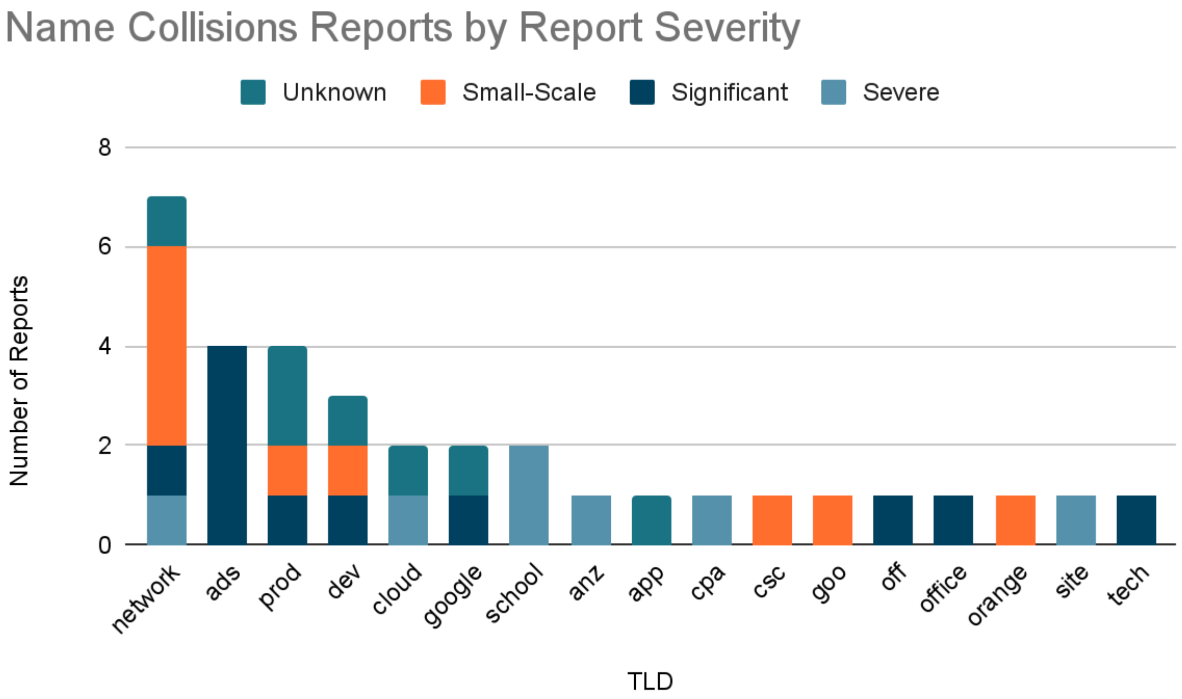
In the following table, we list the count for each category as well as sample comments from each report that led us to categorize them accordingly (except for unknown, for which details were too few to categorize otherwise):

Category	Count	Descriptions
Severe	7	“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	

Our analysis shows that only half (50%) were classified as either severe or significant. However, as noted previously, the text on the submission form suggests that reports are for systems “suffering demonstrably severe harm as a consequence of name collision” and that emergency response actions would be taken “only where there is a reasonable belief that the name collision presents a clear and present danger to human life.” Thus, either our classifications are inaccurate, the reports understate the magnitude of the problems experienced, and/or the reports were submitted notwithstanding the suggested criteria—perhaps in an effort to officially document the problem.

The following figure shows a plot of the severity of the 34 reports by TLD:



Of the reported TLDs, 14 (83%) included at least one report categorized as causing significant or severe impact. Thus, severity was not isolated.

4.4. 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.

4.5. Other Observations

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

First, 8 of the reports mentioned "remote users" or "VPN" (Virtual Private Network). These account for 33% of reports submitted by organizations and 17% of all reports. 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 6), and responses to our survey (see Section 9), 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.

5. Web Search Results Analysis

We now analyze the results of the Web search for “127.0.53.53” (see Section 3.4). Each of these results represents a circumstance in which the IP address 127.0.53.53 was unexpectedly observed in connection with resolving a given domain name ending in a TLD which has been recently introduced into the root zone (with one exception, which will be shown hereafter) as part of the new gTLD program. Thus, we cannot evaluate how often 127.0.53.53 was observed when name collisions were experienced, as this data set *only* includes experiences of name collisions where 127.0.53.53 was observed. However, we again refer the reader to Section 9, where we describe a survey distributed to individuals and organizations *potentially* affected by the delegation of new TLDs, the results of which have no such bias.

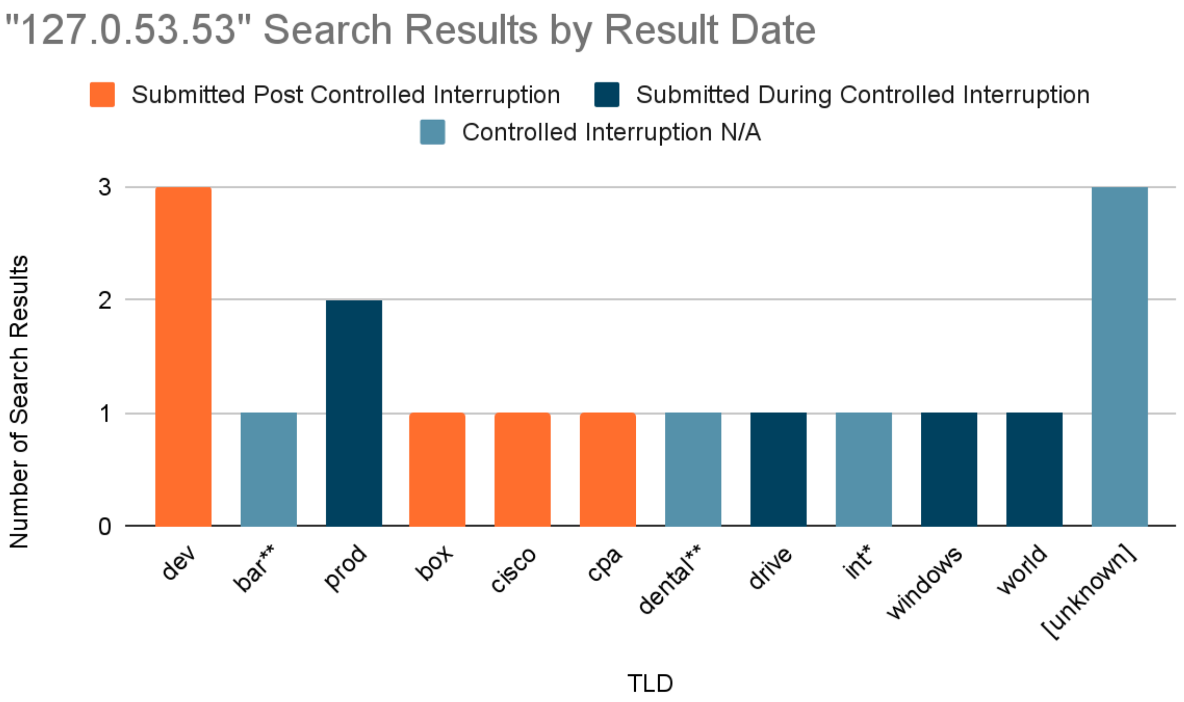
5.1. TLD Statistics

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

Category	**Count**	**Subcat. %**	**Total %**
Total search results	**17**	**100%**	**100%**
do not include TLD	3	18%	18%
include TLD	**14**	**82%**	**82%**
delegated prior to new TLD program*	1	7.1%	5.9%
delegated as part of new TLD program	**13**	**93%**	**76%**
prior to controlled interruption (pre-Aug 2014)**	2	15%	12%
with controlled interruption (Aug 2014 or later)	**11**	**85%**	**65%**
result date is during controlled interruption	5	45%	29%
result date is post controlled interruption	6	55%	35%
Total TLDs in search results	**11**	**100%**	**100%**

delegated prior to new TLD program*	1	9.0%	9.0%
delegated as part of new TLD program	**10**	**91%**	**91%**
prior to controlled interruption (pre-Aug 2014)**	2	20%	18%
with controlled interruption (Aug 2014 or later)	8	80%	73%

The following plot shows the distribution of search results by TLD, including those that were not delegated as part of the new gTLD program (*) and those that were delegated prior to controlled interruption (**). For those results that were associated with the 8 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. As noted previously, the mappings to “127.0.53.53” were observed for the `bar` and `dental` TLDs (both marked with **) even though they are labeled “Controlled Interruption N/A” because they were delegated prior to the start of controlled interruption.



Of the search results corresponding to TLDs delegated as part of the new gTLD program (i.e., excluding int), only 38% were dated during the controlled interruption period for the TLD. These correspond to 45% when only considering the TLDs that were delegated after controlled interruption (i.e., excluding `int`, `bar`, and `dental`). These fractions are comparatively lower than the 74% observed in our analysis of the reports submitted to ICANN (see Section 4.1). However, we note that `dev`, `box`, `cisco`, and `cpa` all continued exhibiting controlled interruption behavior (i.e., returning 127.0.53.53 for non-existent domain names) for 1,042 days, 78 days, 193 days, and (at least) 644 days, respectively, according to DNSDB (see Section 7).

The dates for search results for `dev` and `cpa` were prior to the date that the controlled interruption IP address was last observed. However, the dates of the search results for `box` and `cisco` were past the dates for which the controlled interruption IP address was last observed. Among the possible explanations for the discrepancy are the following. The passive sensors contributing to the historical DNSDB mappings did not have sufficient network placement to observe the controlled interruption experienced by those that posted the report found in the search results. Or it could be that the report (i.e., associated with the Web search result) was made long after controlled interruption was experienced.

The only inexplicable instance of controlled interruption is the one search result corresponding to the `int` TLD. The `int` TLD was delegated in 1988, and we have no data to suggest that it implemented controlled interruption, other than the search result itself.

5.2. Applications In Use

Because the search results often contained more detail than the name collision reports, we were able to glean more about each incident. In 12 (63%) of the 19 reports, a primary application was identified associated with the incident. In cases where the main application was unclear, we categorized it “unknown.” This included cases where we inferred that the application might simply be a diagnostic test but that the main application was something else. The resulting categorization was imperfect but still provided some insight into the use case leading to the collision.

Category	**Count**	**Subcat. %**	**Total %**
Total search results	**17**	**100%**	**100%**
no application identified	7	41%	41%
application identified	**10**	**59%**	**59%**
Web browser	2	20%	12%
ping	2	20%	12%
Apache Kafka (unit testing)	1	10%	5.9%
gitlab-ci-multi-runner	1	10%	5.9%
php, tnsping	1	10%	5.9%
RDP	1	10%	5.9%
SSH	1	10%	5.9%
valet	1	10%	5.9%

First, we note that these results show that there is a variety of applications with which users have experienced name collisions. Additionally, of the search results for which applications were inferred, Web browsers accounted for only 20%.

5.3. Name Collisions Root Causes

The detail in the search results also allows us to better understand the root cause of name collisions affecting applications and end users. We begin with discussion of configurations that contribute to name collisions and then present our findings. Note that these configurations include—but are not limited to—the scenarios described in section 2.3.3 of the NCAP study 1 RFP[132] and section 2.2 of the NCAP study 1 report[133].

Private and Non-private. Much of this document refers to the private use of TLD namespace as the primary cause of name collisions. This is the case in which systems use a presumably non-existent TLD to name resources that they wish to access. If queries for domain names under that TLD reach the public DNS authoritative servers, then there is a name collision. While the private use of TLD namespace seems to be the most prevalent use case for name collisions, there are situations in which name collisions do not involve the private use of domains. We refer to such use as non-private. Name collisions involving non-private use of domain names are typically associated with the use of multi-label, unqualified domain names (discussed hereafter).

Single- and Multi-Label Unqualified. Unqualified names are those that are not intended to be resolved without the application of a DNS suffix (see Section 2.1). There are two variants to those names: single-label (e.g., `foo`) and multi-label (e.g., `foo.bar`). Single-label names traditionally do not resolve to an IP address (exceptions are described later in this section), making them a clear candidate for application of a DNS suffix for proper resolution. In contrast, multi-label, unqualified names have the appearance of being fully qualified, simply because they have more than one label. Yet multi-label, unqualified domain names are known to be used in practice. In the case where the right-most label of a multi-label, unqualified name corresponds to a TLD which has (relatively recently) been delegated is used as the unqualified domain name, the search suffix logic might result in the name being resolved without qualification—ending in a name collision. For example, if `foo.bar` is used as an unqualified domain name, and `bar` is delegated, then `foo.bar` might resolve as if it were fully qualified, regardless of which DNS suffixes are available to be applied.

DNS Suffix Devolution. Some systems use a technique referred to as *DNS suffix devolution* to resolve an unqualified domain name. Given the DNS suffix `foo.bar.com`, suffix devolution involves attempting to resolve the unqualified name `www` first with `www.foo.bar.com` then with `www.bar.com`, etc. An observed variant of this is the following[134]. Given the DNS suffix

[132] https://www.icann.org/en/system/files/files/rfp-ncap-study-1-09jul19-en.pdf

[133] https://www.icann.org/en/system/files/files/ncap-study-1-report-12feb20-en.pdf

[134] https://www.reddit.com/r/sysadmin/comments/2jcdso/workstations_resolving_domainlocal_to_12705353/

`bar.local` and the unqualified name `www`, the system attempts to resolve `www.bar` if `www.bar.local` does not resolve. If `bar` corresponds to a TLD that is newly delegated, then there is a name collision.

Deliberately Unresolvable. One of the causes of name collisions involving unqualified names is that a system or user expects an unqualified name to ultimately resolve in a certain way. In order for this ultimate resolution to work as expected, certain *intermediate* iterations of suffix application (or not) should not resolve. However, in some cases, the user or system uses a name with the expectation that *ultimately* it will not resolve. We refer to these names as *deliberately unresolvable*.

Single-Label Resolution. While the DNS protocol does not prohibit domain names with only a single label (e.g. “dotless domains”) from resolving to an IP address, new gTLDs are administratively prohibited from allowing this type of resolution[135]. Nonetheless, *single-label resolution* has been enabled for at least some gTLDs, and applications take advantage of this functionality. Whether explicitly or inadvertently, this behavior has resulted in name collisions of various types.

Web Search Term. Many Web browsers use a single input area for users to enter either a search string, a domain name, or a URL—any of which must eventually be converted to a URL. Behavior across browsers varies as to the handling of such an input to make this determination. Some browsers attempt to resolve a single “word” (i.e., no spaces) as a domain name, only using it as a search term after it has been shown to not resolve. In such cases, a word intended as a search term that corresponds to a TLD that has been delegated and configured for single-label resolution, results in a name collision.

VPN. As discussed in Section 4.5, name collisions are often manifest when a VPN is in use. In such cases, the system is potentially operating under two network environments, and what might otherwise be *controlled* use of private namespace can be exposed to public authoritative DNS, resulting in name collisions.

We categorize the use cases according to the following:

- *Private Namespace.* Was the user’s system using the TLD in a private context?
- *Qualification.*
    - Was an unqualified single- or multi-label name the target of resolution? For unqualified, single-label names, was some form of suffix devolution used for search list processing? For unqualified, multi-label names, was the name non-private?
    - Was the name fully qualified? If so, was the name intended to be deliberately unresolvable?

[135] https://www.icann.org/en/announcements/details/new-gtld-dotless-domain-names-prohibited-30-8-2013-en

  - Was a single label being resolved to an address? Was the intention for the domain name to be used as a Web search term?
- *VPN.* Was a VPN involved? If so, was the name private?

Category	**Count**	**Subcat. %**	**Total %**
Total	17	100%	100%
Private Namespace			
Private	11	65%	65%
Non-Private	2	12%	12%
N/A	1	5.9%	5.9%
Unknown	3	18%	18%
Qualification			
Unqualified	**6**	**35%**	**35%**
Single-Label	**4**	**67%**	**24%**
Suffix Devolution	1	25%	5.9%
Multi-Label	**2**	**33%**	**12%**
Non-Private	2	100%	12%
Fully-Qualified	**10**	**59%**	**59%**
Deliberately Unresolvable	3	30%	18%
Single-Label Resolution	**1**	**5.9%**	**5.9%**
Search Term	1	100%	5.9%
Unknown	2	12%	12%
VPN	**2**	**12%**	**12%**
Private	1	50%	5.9%
Unknown	1	50%	5.9%

Perhaps the most interesting aspect of the analysis is that the causes are so diverse, particularly for such a relatively small dataset. Nearly every conceivable use case is represented. As mentioned previously, use of private namespace accounts for the majority (65%) of search results. The use of unqualified names with search list processing accounted for only 35% of cases, with two thirds of those involving single-label, unqualified names and the rest involving multi-label, unqualified names. Of the nearly 60% of cases that involved fully-qualified domain names, 30% were cases where the fully-qualified name was ultimately not intended to resolve. In 12% of cases VPN usage was mentioned—compared to 17% reported in the name collision reports submitted to ICANN. There was one case of nuanced DNS suffix devolution. Finally, there was one case where single-label resolution was at play, and it corresponded to the use of a label as a search term.

5.4. Other Observations

Among the other observations were the following. First, while all 17 search results contained a reference to the controlled interruption IP address, 127.0.53.53, 13 (76%) of those additionally included a reference to ICANN and controlled interruption; only 4 (24%) did not reference ICANN. Thus, there was a relatively high success rate in associating the IP address 127.0.53.53 to ICANN and controlled interruption—for those that observed the IP address.

Second, we note the sentiment expressed in each of the scenarios gleaned from search results was generally neutral (16 results or 94%). That is to say that the public commentary accompanying the situations in which users encountered name collisions was neither positive nor negative towards controlled interruption. In only one instance (6%) did the language convey anger—which was towards both ICANN and Google, the registry for the TLD in question.

6. 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 2.1). 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.

6.1. 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` 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 in a mapping. Thus, we only considered mappings for which the "first seen" timestamp equals the "last seen" timestamp, i.e., it was only observed once.

At this point, we grouped the mappings observed within a timestamp by common suffix of the qname—defined as everything to the right of the first (i.e., left-most) label. We then applied the following additional criteria:

- **Qnames with common suffixes found in groups of three.** Only suffixes found in groups of three were considered, i.e., corresponding to the number of probing queries issued by Chrome.
- **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.

As an example, suppose the following queries were observed:

First seen	Last seen	Query (qname/type)	Reason for Disqualification
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`	More than 1 second
1649687020	1649687020	`polkuhadev.bar.example.com/A`	Group of 2 qnames
1649687020	1649687020	`fvqiyjas.bar.example.com/A`	Group of 2 qnames
1649687020	1649687020	`hnsjmirc.baz.example.com/A`	Group of 1 qname

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

An analysis of the Chrome identification methodology is found in Section 6.4.

Even with the measures we took, there still might be room for false positives. In Section 8, we further filter the suffixes to increase confidence in the data set used for our later analysis.

6.2. Suffix Identification Using WPAD and ISATAP DNS Queries

To identify suffixes using DNS queries related to WPAD and ISATAP, we identified all qnames whose 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.”

We validate our methodology related to DNS suffix identification in Section 6.5.

6.3. Results

6.3.1. Validation of Identification Methods

The total number of DNS suffixes identified in the DNSDB-CI data set was 2,762. The following table shows the counts and percentages of DNS suffixes identified using different combinations of the methods:

Identification Method(s)	Suffixes Identified		
	Count	**Subcategory %**	**Total %**
Chrome, WPAD, *or* ISATAP - *Any*	**2,762**	**100%**	**100%**
Chrome, WPAD, *and* ISATAP - *All*	1,064	39%	39%
Chrome	**1429**	**52%**	**52%**
Chrome *only*	197	14%	7%
Chrome *and* WPAD or ISATAP	1,232	86%	45%
WPAD	**2,084**	**75%**	**75%**
WPAD *only*	360	17%	13%
WPAD *and* ISATAP or Chrome	1,724	83%	62%
ISATAP	**2,065**	**75%**	**75%**
ISATAP *only*	453	22%	16%
ISATAP *and* Chrome or WPAD	1,612	78%	58%

Each percentage in the “Subcategory %” column is taken from the “Count” in the “parent” category or subcategory (i.e., the bolded count most immediately above). The percentages in the “Total %” column are taken from the “Count” in the “Chrome, WPAD or ISATAP - Any” category.

Each method resulted in the identification of between 52% (Chrome) and 75% (WPAD and ISATAP) of all 2,762 suffixes. These percentages show that each identification method contributed to the set of DNS suffixes. To further validate the suffixes identified with each method, we further analyze the contributions of each subsequently.

The subcategories whose label includes “and” (e.g., “Chrome *and* WPAD or ISATAP”) show how many of the suffixes identified by *one* method (e.g., Chrome) were identified by at least one *other* method (e.g., WPAD or ISATAP). Higher values indicate more confidence in the method, i.e., because multiple applications were used in the environment exposing this DNS suffix. For all three methods, the percentage of suffixes identified by at least one other method was at least 45%.

The subcategories labeled “only” (e.g., “WPAD only”) identify the *individual contributions* of each method—that is, how many of the suffixes were identified *only* because the listed method was employed. Larger numbers are a possible indicator that the suffix identification method was inaccurate, finding many suffixes that were not found by any other methodology. However, we also would not expect a zero value because of the diversity of application deployment within network environments. In every case, these figures are under 20% of the total. 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.

6.3.2. Distribution of Suffixes Across TLDs

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 disproportionately high 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 <u>underlined</u> indicate a 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.

6.4. Analysis of Chrome Identification Methodology

We previously identified rules for detecting DNS suffixes by recognizing qnames associated with Chrome browser NXDOMAIN probing behavior (see Section 6.1). Two of the criteria for considering potential suffixes using the Chrome method were that they showed up in groups of three at a given timestamp and that the exact group of three qnames does not show up at any other timestamp. We now mention some statistics with regard to those which were “rejected” from candidacy because of failure to meet that criteria. A total of 21,768 *potential* suffixes were identified before considering the number of mappings with a given suffix at a given timestamp and uniqueness of groups of qnames. Of those 20,336 (93%) were eliminated because they were not in groups of three, and an additional 3 were eliminated because the same qnames were found at a different timestamp. This is a fairly high percentage, and we suspect that some of these are false negatives. However, the intent was to reduce false *positives*. As mentioned previously, 86% of the DNS suffixes identified with Chrome browser identification were also identified by either the ISATAP or the WPAD methodology, and only 14% were found exclusively using the Chrome technique. This percentage is comparable to those of the WPAD and ISATAP methods, which were 17% and 22%, respectively. These numbers provide confidence in the methodology. While a more rigorous validation of this and other methods is possible, it is beyond the scope of the work.

6.5. Validation of WPAD Identification Methodology

As mentioned previously (Section 6.2), 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.

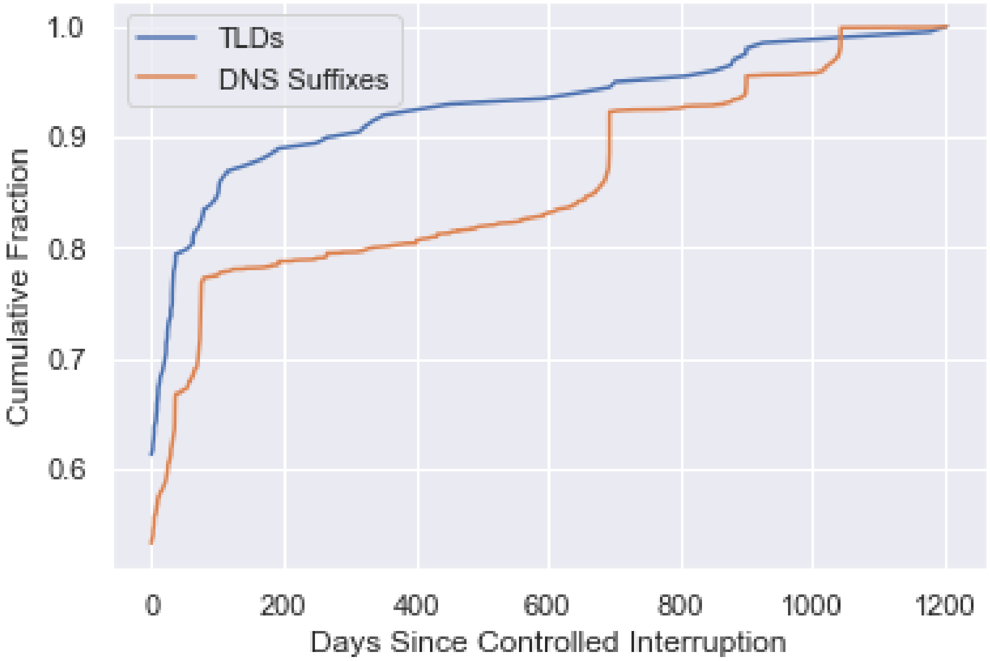
7. Controlled Interruption Analysis

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 8), we reduced the number of DNS suffixes to 2,300, within 200 TLDs—instead of the full set of 2,762 suffixes within 498 TLDs. With this reduced data set we 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.

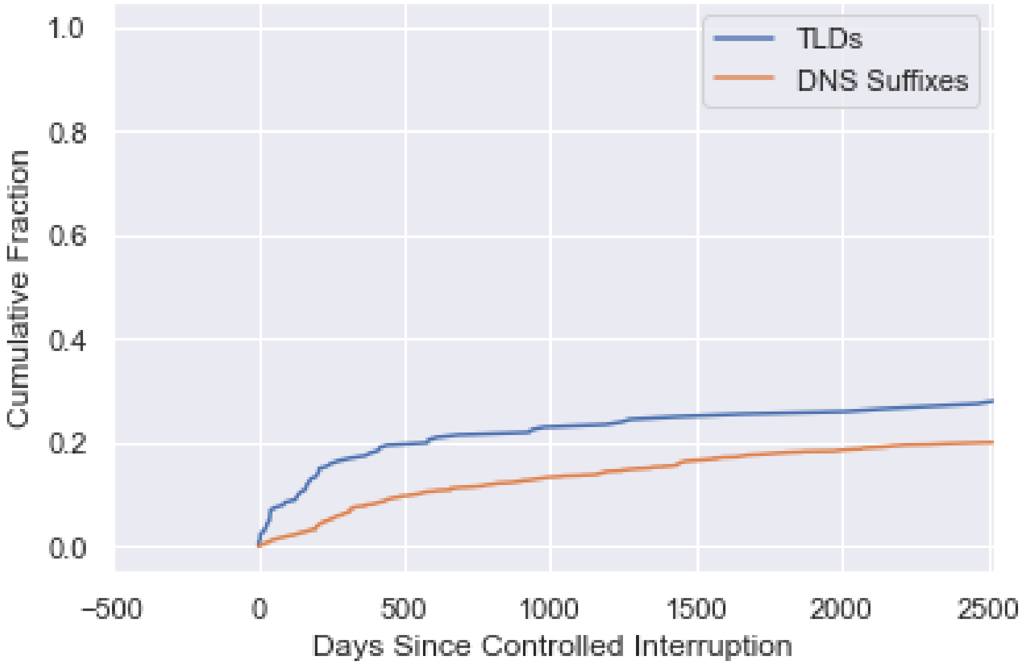
As mentioned previously (Section 2.2), the IP address 127.0.53.53 is returned for all names under a TLD during the first 100 days of its delegation, i.e., the controlled interruption period. By analyzing the mappings in DNSDB-PostCI, we were able to determine how long controlled interruption was observed 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 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 Section 2.4 and Section 2.5). Fortunately, we found no mappings for such qnames in the DNSDB-Post-CI data.

8. Root Server Query Analysis

The DNS suffixes identified in Section 6 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.

8.1. Data Set

We used the DITL data from 2014 through 2021 (see Section 3.3) 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.

Eliminate TLDs. First, we reduced the suffixes by eliminating those that were themselves TLDs. For example, `office` is a TLD, but it was *also* identified as a DNS suffix through one or more of the identification methods. Thus, DNS queries associated with the suffix `office` because it was a TLD. The rationale behind excluding TLDs was two-fold. First, by including a TLD, our filter would include *all* queries ending with that TLD. Many of those queries would be false positives, and we have no way to reliably exclude false positives from the data set when the suffix is a TLD. Additionally, as mentioned previously, one of the objectives of this analysis is to identify organizations from which the DNS suffix originated, as part of root cause, by using the suffix itself. For example, the suffix `acme.network` originating from a network with name “ACME” would support an association between the network and the DNS suffix. However, a single label is typically too generic to help us associate suffixes to organizations in that way.

Further Filtered TLDs. 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.

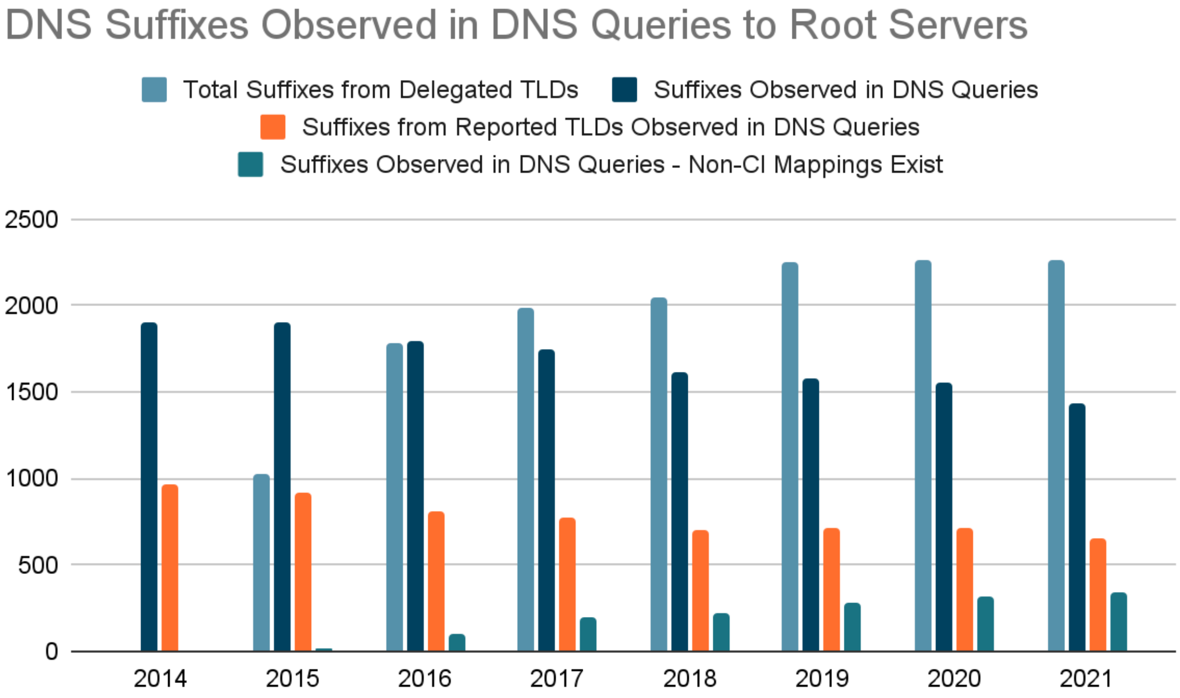
	Suffixes	**TLDs**
All DNS Suffixes	2,762	498
DNS Suffixes - no TLDs	2,300	200
DNS Suffixes - no TLDs and further filtered (TLD has at least one WPAD suffix, one ISATAP suffix, or more than 1 total suffix)	2,266	166

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 4). 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.

8.2. Results

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 4); 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).

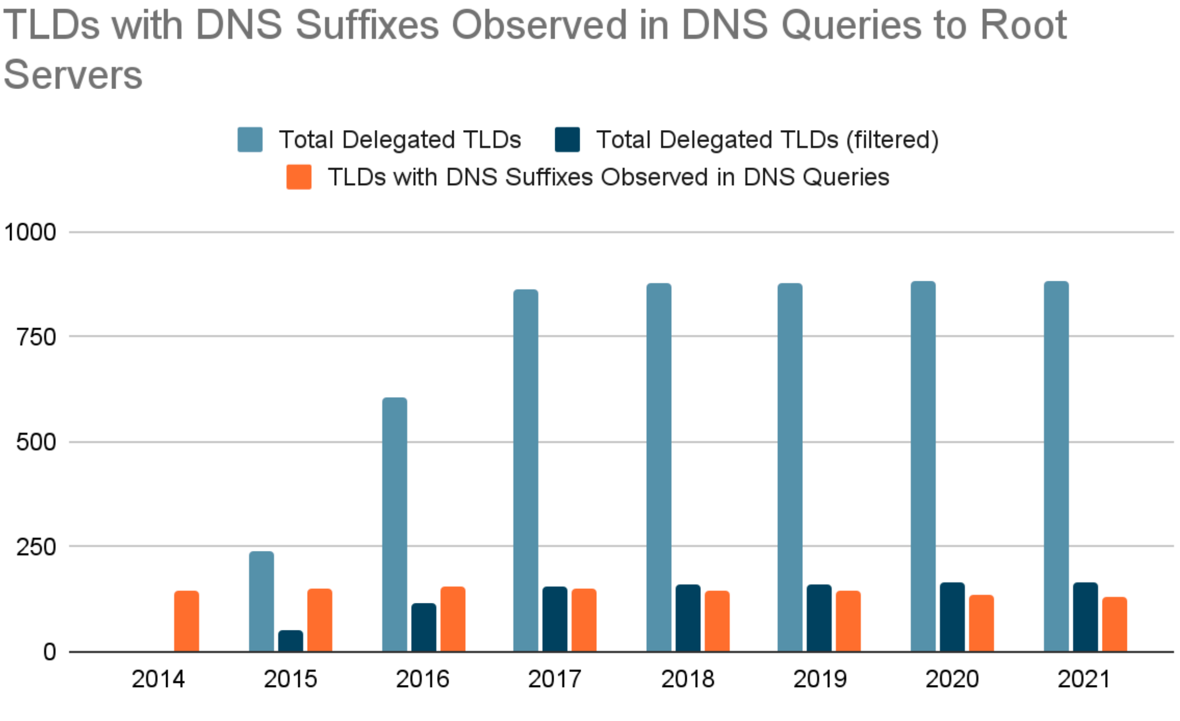


While over 1,900 (84%) of the 2,266 DNS suffixes were observed as early as 2014, the number of suffixes observed in DNS queries has consistently decreased over time, as new TLDs have been delegated, such that in 2021 1,434 (63%) suffixes were observed. Nearly half of those DNS suffixes are associated 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 6 for more.

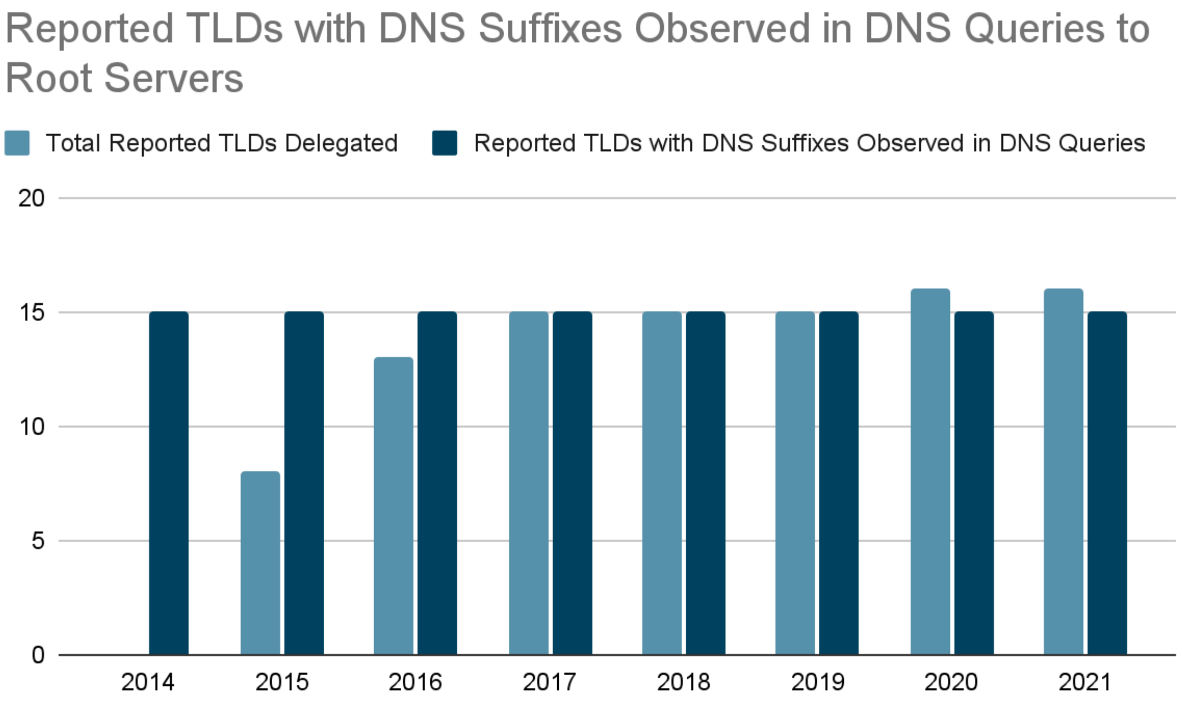
Between 2015 and 2021, there is a steadily increasing number of DNS suffixes observed in query data for which non-CI mappings exist (see Section 7). In 2021, queries were observed for 336 suffixes that had a non-CI mapping. That accounts for 23% of all DNS suffixes observed in queries at the DNS root and15% of all 2,266 DNS suffixes. 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.



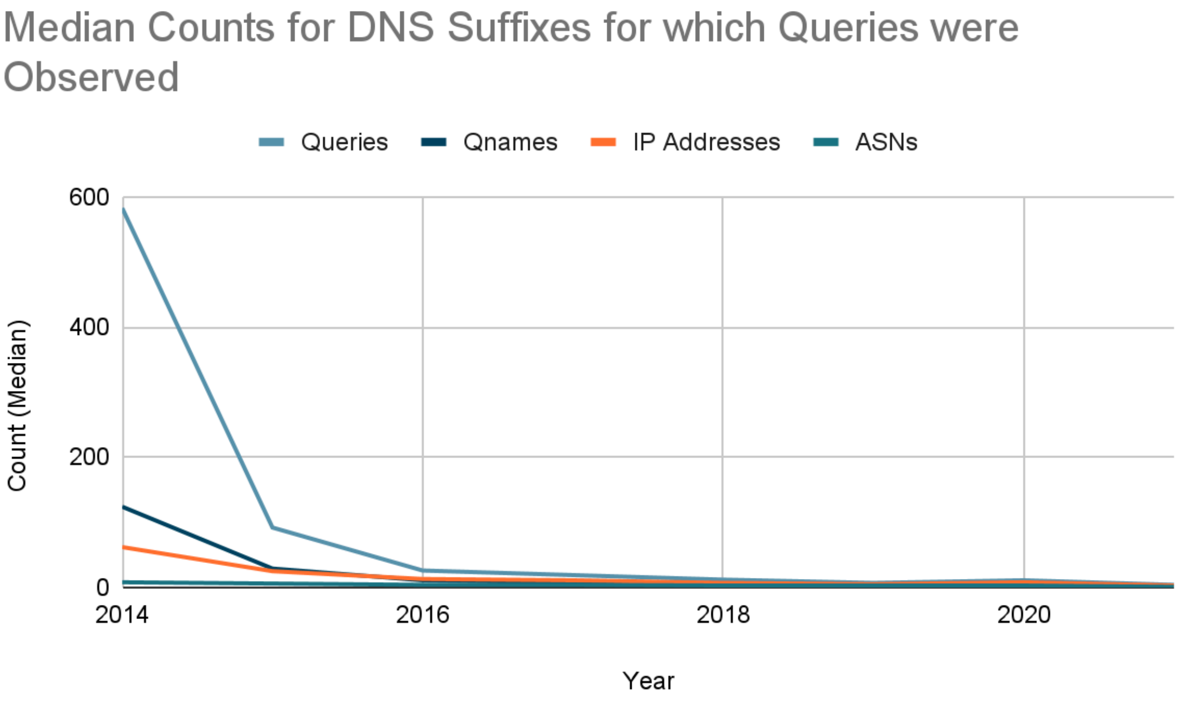
The number of TLDs experiencing name collisions, by observation, has remained relatively steady from 2014, when queries for DNS suffixes associated with 146 TLDs (88% of filtered, 16% of all TLDs) were observed, through 2021, when 133 TLDS exhibited name collision behavior (80% of filtered, 15% of all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

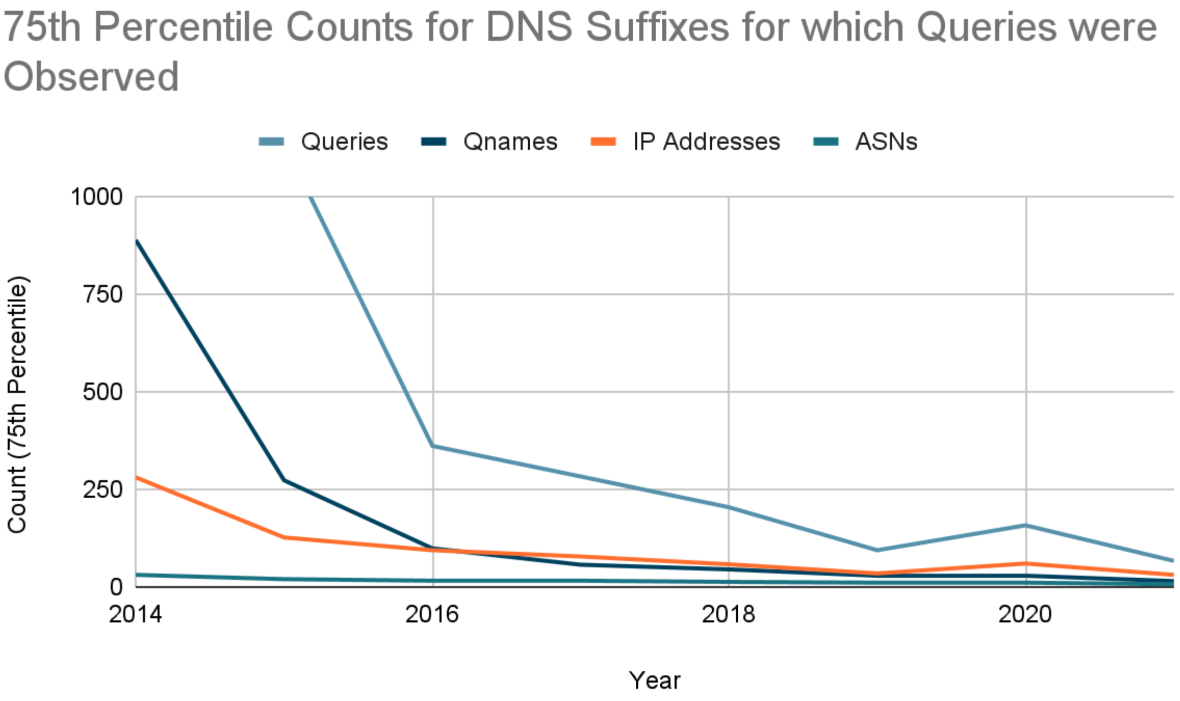


This shows that in every DITL collection between 2014 and 2021, queries for DNS suffixes within 15 (94%) of the 16 reported TLDs, after filtering, were consistently observed. Only the TLD `google` was not observed. While the general trend was mostly consistent, this trend was completely consistent.

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:

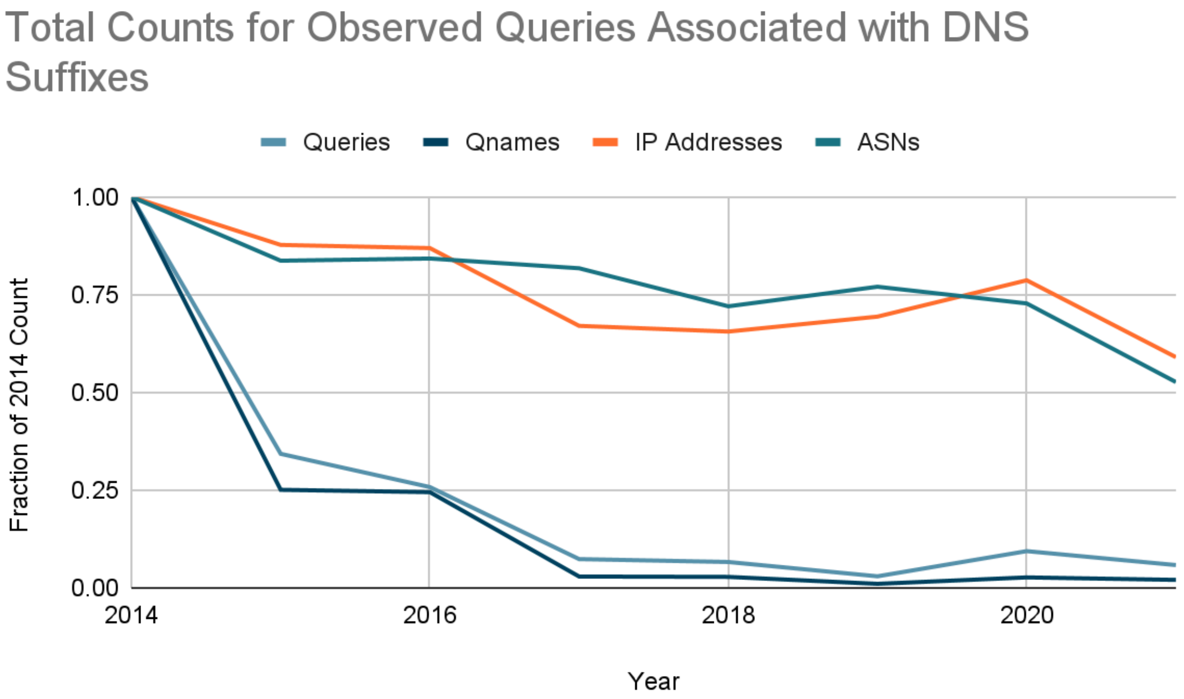


Median Counts for DNS Suffixes for which Queries were Observed



75th Percentile Counts for DNS Suffixes for which Queries were Observed

In the next figure, we show overall counts for DNS queries associated with identified DNS suffixes, as a fraction of those observed in 2014:


Total Counts for Observed Queries Associated with DNS Suffixes

The plot is normalized because of the significant difference in scale between the different categories. For reference, the following table shows the raw counts:

Year	Queries	Qnames	IP Addresses	ASNs
2014	62,305,672	25,937,776	112,374	12,296
2015	21,358,020	6,504,348	98,555	10,287
2016	16,061,683	6,349,761	97,640	10,356
2017	4,586,613	754,204	75,294	10,050
2018	4,126,353	729,336	73,658	8,854
2019	1,846,412	268,356	77,951	9,469
2020	5,855,426	695,784	88,393	8,944
2021	3,636,318	531,233	66,304	6,472

In all plots, a clear trend of decreasing per-suffix and overall 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[136] or qname minimization[137], which we explain in the following paragraphs.

[136] RFC 8806
[137] RFC 7816

The local root specification was first published in November 2015 and updated in June 2020. It provides guidance for serving a copy of the root zone on a recursive resolver. This keeps the resolver from having to issue any queries to the root servers because it has all the answers it needs locally. It thus achieves benefits of both privacy and performance. There are currently no research studies to provide insight into the prevalence of local root deployment. However, the publication date of the original specification for local root deployments was *after* the prominent decrease in per-suffix and total counts related to name collisions, which was first observed in April 2015.

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, qname-minimizing resolvers take advantage of the fact that the only *required* component is `com`, i.e., to elicit a referral. They use various techniques to conceal more specific query information from authoritative servers. Recent studies suggest that qname minimization affects 12% of Internet resolvers and 40–48% of queries as of 2018[138]. We consider the effects of qname minimization in Section 8.3.

To gain additional insight into the causes of the query behaviors we observed, we supplement our quantitative measurements with a qualitative study, which we discuss in Section 9.

8.3. Qname Minimization Considerations

The data that has been presented thus far has been compiled independent of qname minimization. However, because qname minimization has seen an increase in deployment, and its effects might contribute to some of the downward trends in our analysis, we now perform additional analysis that takes qname minimization into account.

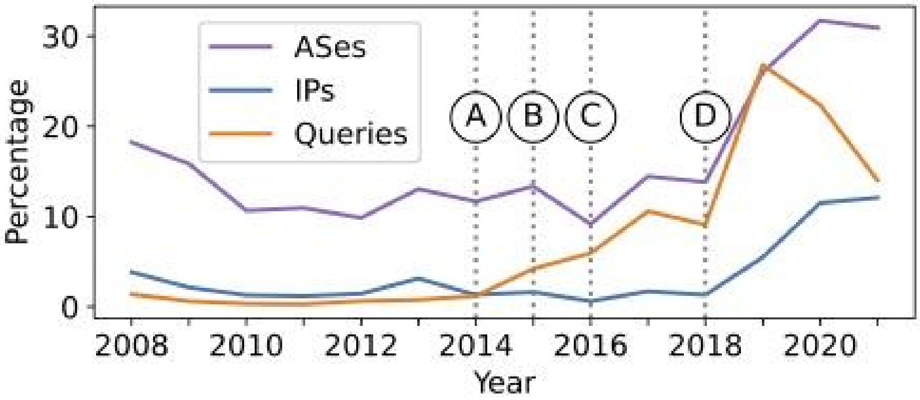
8.3.1. Summary of Recent Study of Qname Minimization

We first summarize recent work analyzing the deployment of qname minimization on resolvers that queried A-root during between the years of 2008 and 2021, using the yearly DITL collection as its data source[139]. In that work, the process for determining qname minimization behavior was as follows. A resolver was evaluated for qname minimization by testing for the following two query behaviors during the collection period: 1) the resolver issued a minimum of five queries for qnames other than the root name; and 2) the resolver issued no query with a qname having more than one label. If a resolver met both requirements, then it was considered to be qname-minimizing. If it met only the first, then it was considered to be non-qname-minimizing. If it met neither requirement, then no assessment could be made.

[138] https://www.nlnetlabs.nl/downloads/publications/devries2019.pdf

[139] “Fourteen Years in the Life: A Root Server's Perspective on DNS Resolver Security” by Alden Hilton, Casey Deccio, and Jacob Davis. To appear in *Proceedings of USENIX Security ’23*.

We include below the plot from that work that shows the percentage of IP addresses (of the subset that could be evaluated, based on the five-query minimum) that exhibited qname-minimizing behavior:



Also shown are the percentage of ASes for which at least one qname-minimizing resolver was observed and the percentage of queries corresponding to the qname-minimizing resolvers. The labeled vertical lines represent (A) the submission of the initial qname minimization Internet Draft, (B) its adoption by the unbound resolver, (C) its adoption by Knot resolver and its publication as an RFC, and (D) its adoption by BIND resolver. From the vantage point of A-root, the percentage of resolvers that use qname minimization has increased from 1% to 12% between 2018 and 2021. The percentage of overall queries that come from qname-minimizing resolvers has risen from 1% to 14% between 2014 and 2021, with it reaching as high as 27% in 2019.

Notably, the upward trend in deployment of qname minimization does not correlate with the downward trend associated with the name collision queries observed at the root servers. While significant uptick of qname-minimizing resolvers did not occur until 2019 with its inclusion in BIND, the significant decrease in per-suffix name collision queries occurred in 2015, which was the first DITL collection after controlled interruption was instituted.

8.3.2. Application of Qname Minimization Data

We next sought to isolate the resolvers identified as non-qname-minimizing and run our analysis again on *only* those, so we could compare the trends observed in this latest analysis with those resulting from the analysis that did not consider qname minimization (i.e., from Section 7.2). However, there were three challenges with this. *First*, the IP addresses observed at A-root constituted only 40% of all IP addresses seen at the collective root servers (except I-root and L-root, which anonymize their IP address data) during the 2021 DITL collection. Even so, these IP addresses represented 95% of ASes from which queries were received by the collective root servers. *Second*, only 36% of the IP addresses querying A-root met the criteria for qname minimization evaluation in 2021, corresponding to 15% of the total IP addresses observed in 2021. Of those, 88% of IP addresses exhibited behavior characteristic of non-qname-minimizing resolvers. Thus, the percentage of 2021 IP addresses that are used for our analysis is 13%.

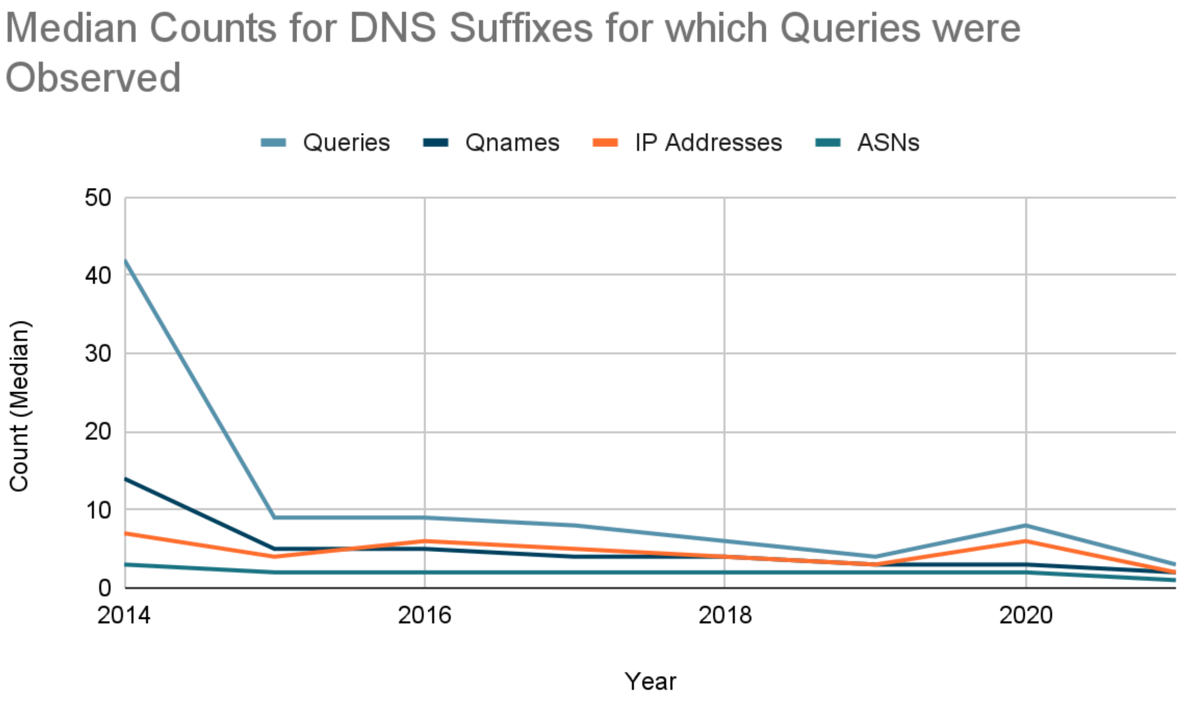
Finally, the set of IP addresses observed in DITL collections prior to 2021 is not the same as the set observed in 2021; various factors over time contributed to the variance between those sets.

We used the IP addresses of the non-qname-minimizing resolvers identified in the 2021 DITL collection as the basis for carrying out the analysis in the previous years. We did this under the assumption that if a resolver was not using qname minimization in 2021, then it was likely not using qname minimization before 2021. This assumption greatly simplified the data set we were working with and its analysis. A summary of the numbers of IP addresses comprising the analysis for each year since 2018 is found in the following table. In each case, the percentage reflects the percentage of all IP addresses observed in the given DITL collection year:

DITL Year	IP Addresses (all root servers)	IP Addresses (only A-root 2021)	Qname Min. Evaluated	Non-Qname Min.
2018	17,017,222	7,047,980 (41%)	1,205,290 (7%)	1,121,513 (7%)
2019	12,651,567	7,071,314 (56%)	1,511,110 (12%)	1,395,088 (11%)
2020	17,343,285	8,718,048 (50%)	2,089,481 (12%)	1,893,877 (11%)
2021	26,463,953	10,612,429 (50%)	3,845,577 (15%)	3,380,341 (13%)

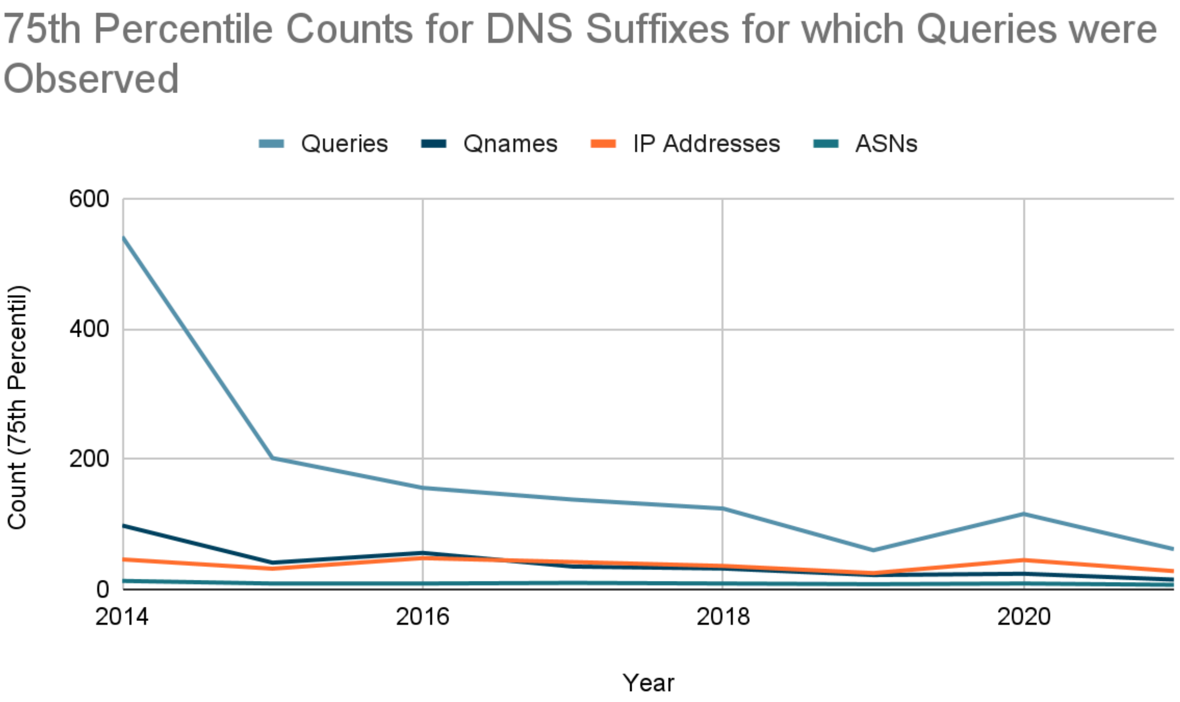
Thus, the sample of data from which we take our analysis ranges from 7% (2018) to 13% (2021). Sample data prior to 2018 is not currently available.

The median per-suffix counts for queries, unique qnames, IP addresses, and ASNs is shown in the following figure:

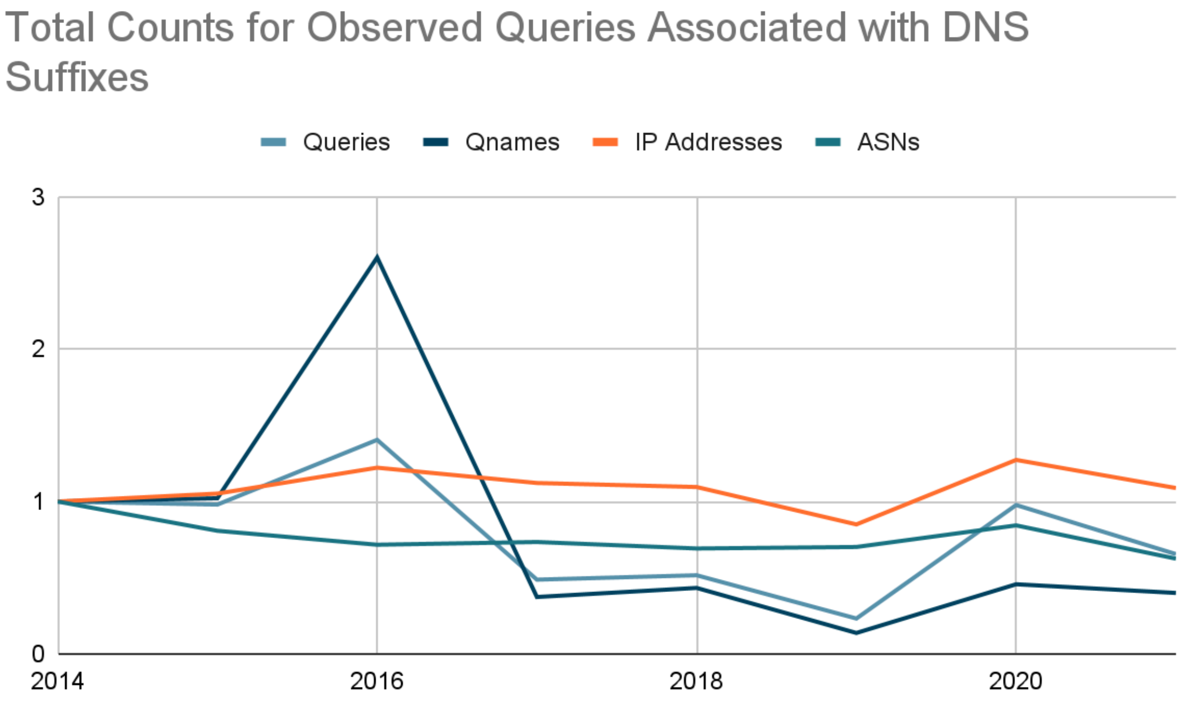


The raw numbers are, expectedly, much lower each year in this plot than they are in the previous plot, which considers the entire set of querying IP addresses; this is due to the very fact that we are working with only a subset of the data. However, the trends in this plot match those in the previous plot, especially in the following ways: 1) both plots show a significant decrease in median counts between 2014 and 2015; both plots show relatively little change between 2015 and 2021; and 3) both plots show a slight increase in median counts in 2020.

The trends associated with the per-suffix 75th percentile counts for non-qname-minimizing resolvers also match those of the plots that consider all resolvers:



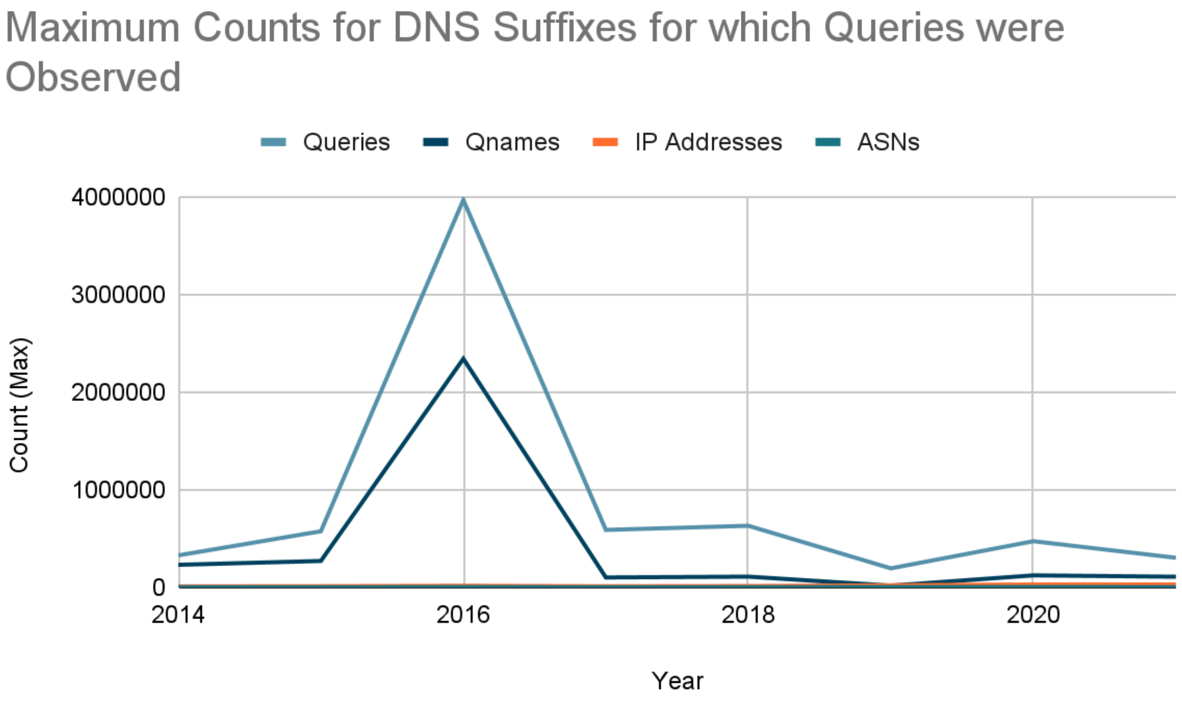
Finally, we consider the total counts associated with name collision queries, across all DNS suffixes, each shown as a percentage of the respective 2014 value:



The corresponding raw numbers are shown in the following table:

Year	Queries	Qnames	IP Addresses	ASNs
2014	5,439,148	1,315,878	172,559	45,634
2015	5,330,047	1,344,184	181,541	36,869
2016	7,639,202	3,423,237	210,865	32,692
2017	2,650,592	491,464	193,517	33,514
2018	2,807,826	569,824	188,971	31,571
2019	1,263,280	181,099	146,644	32,020
2020	5,314,238	601,583	219,667	38,490
2021	3,562,760	526,523	187,830	28,524

In this case, a consistent trend is hard to observe within the data itself, and it differs significantly from its counterpart, which includes queries from all IP addresses, rather than just non-qname-minimizing IP addresses. Because the median is so consistent, these relatively high and inconsistent counts are likely related to outliers—suffixes that receive *many* more queries than the median or 75th percentile, within the non-qname-minimizing IP addresses. To test this, we plot the maximum count values across all DNS suffixes, for each of the years:



There are features in the plot that clearly demonstrate outlier behavior, the most prominent of which is the relatively high number of queries and unique qnames queried for name collisions in 2016. The cause of these outliers requires further research, but it is outside the scope of this work. It is sufficient to indicate that outlier behavior is at play with the plot showing the total counts associated with name collisions queries.

Based on the analysis presented herein, we conclude that the trends related to name collision DNS queries observed at the root servers from DITL collection data are not affected by qname minimization behaviors.

9. 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.

9.1. 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 Appendices C and D, respectively.

9.2. 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 text of the message is in Appendix E.

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 (see Section 8).
- 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. The text of the message is in Appendix F.

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 4 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 4. However, as will be noted, this bias has little impact on our findings because the response rate was so low.

9.3. 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.

9.3.1. 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.

9.3.2. 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.

9.3.2.1. 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.

9.3.2.2. 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.

9.3.2.3. 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.

9.3.2.4. 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.

9.4. 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.

10. 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.

10.1. Findings

Private use of DNS suffixes is widespread. It is clear from the data that private use of DNS suffixes is not isolated. Apparently private use of DNS suffixes is exhibited within over half of newly delegated TLDs, even though 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 are supported strongly by 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 colliding with newly delegated TLDs has decreased 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/or 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. Some anecdotal data submitted by survey respondents supports the evidence of that decrease. We also reference outside studies that show some uptake of qname minimization, which reduces the query context available at root servers (see Section 8).

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.** Of the survey respondents that indicated that they used of TLDs, 70% reported having experienced technical issues related to their suffix. Of those, 43% experienced the problems within days of delegation of the TLD. Over two-thirds (71%) of organizations experiencing technical problems indicated that they knew that the issues were related to TLD delegation before the problem was resolved. It appears that most of the ineffectiveness was due to the controlled interruption IP address not even being observed, which occurred in 71% of cases, according to the survey. However, when the controlled interruption IP address *was* observed, the success rate in identifying ICANN and controlled interruption as the cause was between 50% and 76%, according to the survey results and the Web search results analysis, respectively.

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 collision reports, half (17 or 50%) of the reports imply severe or significant impact to the reporting entities.

The public response to controlled interruption was overall neutral. Name collisions and controlled interruption certainly impacted various individuals and organizations. Nonetheless in forums where users or administrators publicly posted questions or experiences with controlled interruption, the overall sentiment was neither positive nor negative, but neutral. **Evidences.** A sentiment analysis of the Web search results revealed that in 94% of cases, neither positive nor negative feelings were expressed towards controlled interruption. In only one case (6%) was negative sentiment expressed.

Name collisions were diverse, both in terms of the application involved and their root causes. Multiple applications were involved with name collisions, some with which users interface directly and some which are more process-driven. Name collisions were caused by the use of both private and non-private namespace. They were caused by the use of domain names that were fully-qualified and unqualified, including unqualified names with single and those with multiple labels. **Evidences.** Eight different applications were responsible for the 10 Web search results that revealed an application affected by name collisions. No single application was responsible for more then 20%, including Web browsers. While nearly two thirds (61%) of collisions identified in the Web search results were caused by the private use of TLD namespace, 10% involved the use of namespaces that were non-private. The Web search results also showed that name collisions were encountered in cases where a name was fully-qualified (59%), unqualified (35%), and even where a single-label was used (5.9%). Additionally, the use of unqualified domain names involved both single-level (67%) and multi-label (33%) unqualified domain names.

10.2. 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.

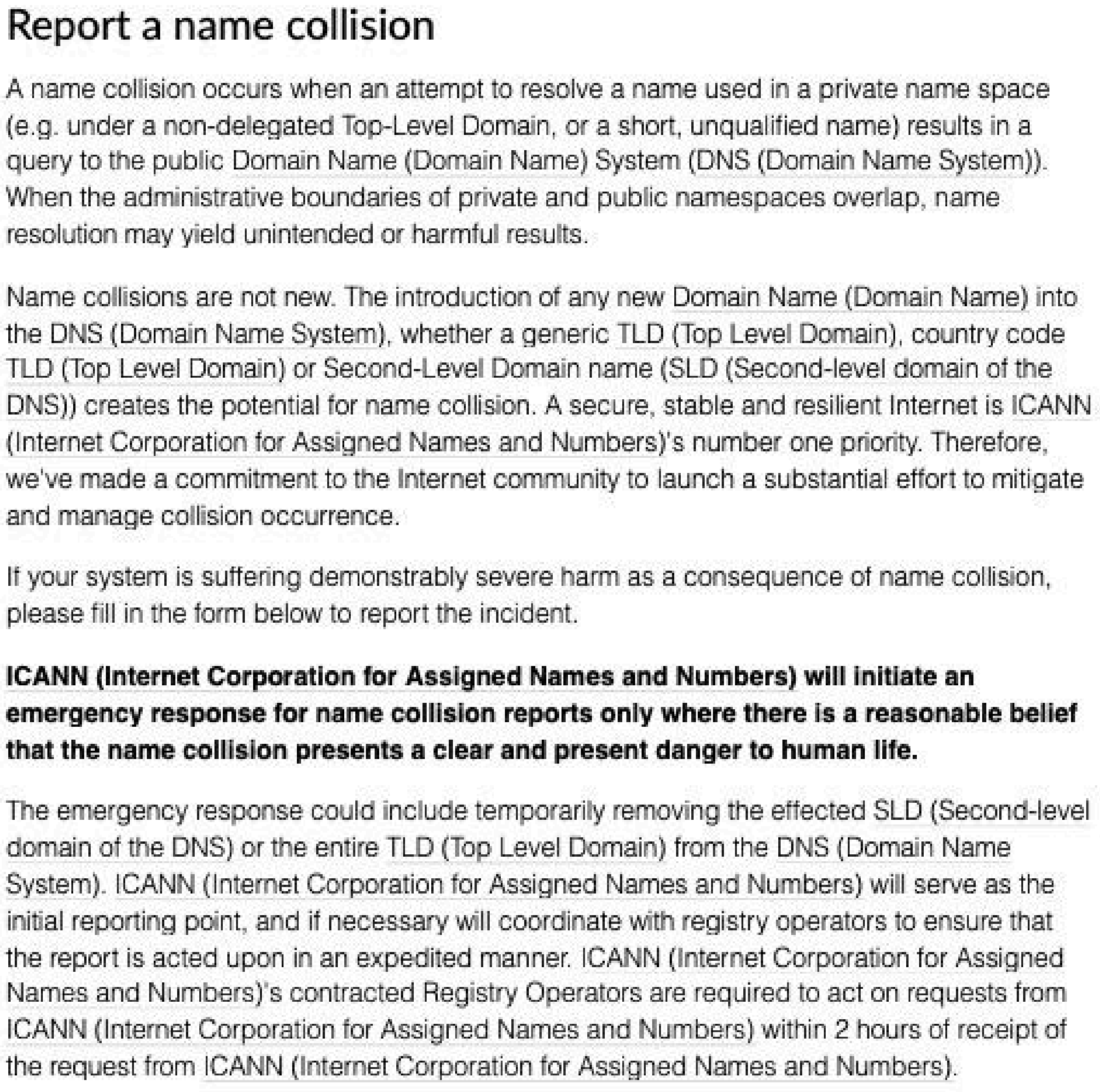
To fill the knowledge gap on the experiences of organizations, we propose additional work, targeting *analysis* and *reach-out* related to the suffix-ASN mappings. The goal in both of these is to better understand how DNS suffixes are being used and to further our understanding of organizational impact 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. Evidence of mobile devices was observed in both root server queries and from name collision reports submitted to ICANN. 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. Further investigating the use of private DNS suffixes on mobile devices will not only help us better understand the configuration trends of mobile devices but might also help us more accurately determine the cause(s) of decreasing DNS queries for private-use suffixes over time.
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`, by their very nature, are 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 - Name Collisions Report Form



Report a name collision

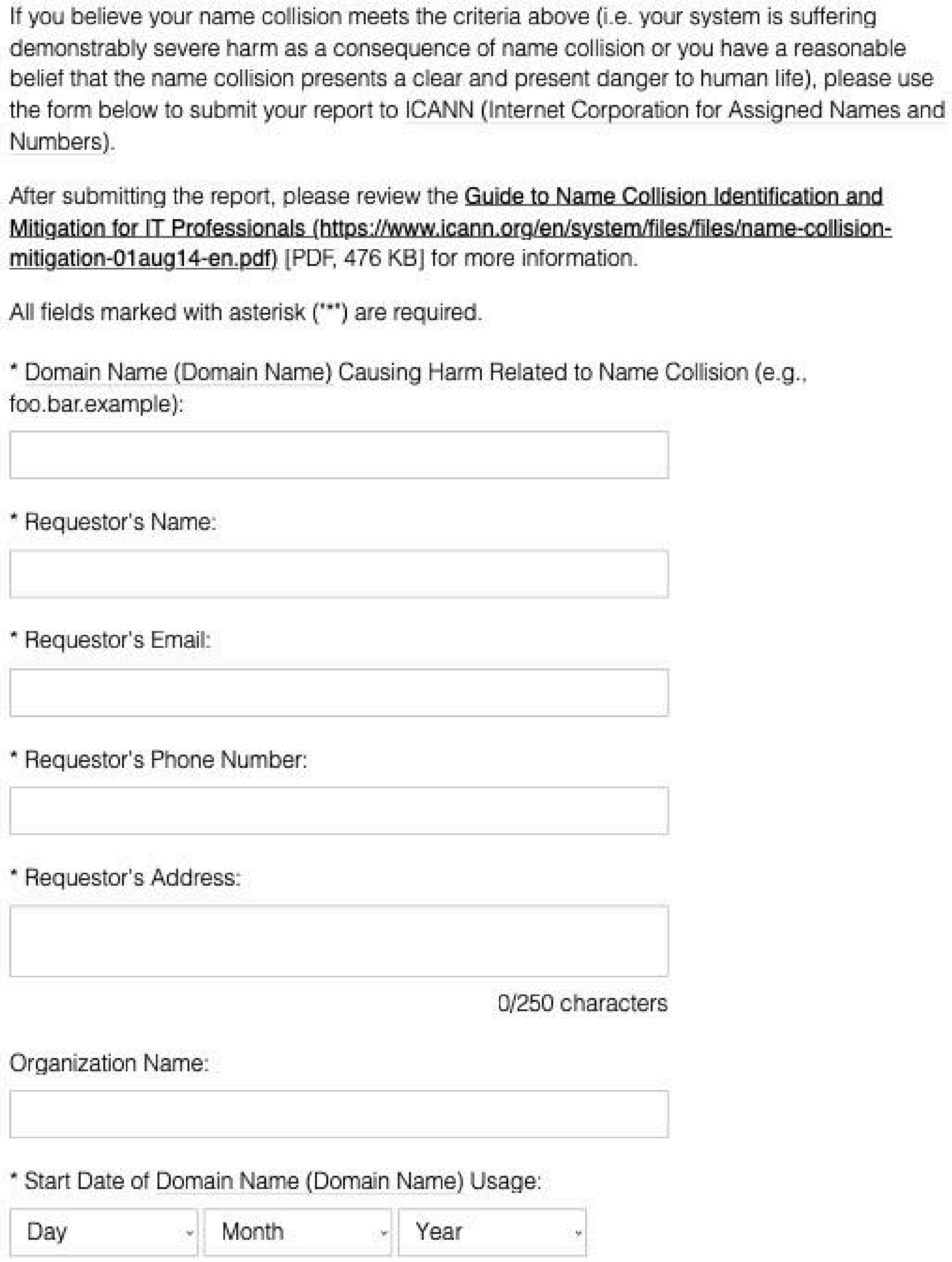
A name collision occurs when an attempt to resolve a name used in a private name space (e.g. under a non-delegated Top-Level Domain, or a short, unqualified name) results in a query to the public Domain Name (Domain Name) System (DNS (Domain Name System)). When the administrative boundaries of private and public namespaces overlap, name resolution may yield unintended or harmful results.

Name collisions are not new. The introduction of any new Domain Name (Domain Name) into the DNS (Domain Name System), whether a generic TLD (Top Level Domain), country code TLD (Top Level Domain) or Second-Level Domain name (SLD (Second-level domain of the DNS)) creates the potential for name collision. A secure, stable and resilient Internet is ICANN (Internet Corporation for Assigned Names and Numbers)'s number one priority. Therefore, we've made a commitment to the Internet community to launch a substantial effort to mitigate and manage collision occurrence.

If your system is suffering demonstrably severe harm as a consequence of name collision, please fill in the form below to report the incident.

ICANN (Internet Corporation for Assigned Names and Numbers) will initiate an emergency response for name collision reports only where there is a reasonable belief that the name collision presents a clear and present danger to human life.

The emergency response could include temporarily removing the effected SLD (Second-level domain of the DNS) or the entire TLD (Top Level Domain) from the DNS (Domain Name System). ICANN (Internet Corporation for Assigned Names and Numbers) will serve as the initial reporting point, and if necessary will coordinate with registry operators to ensure that the report is acted upon in an expedited manner. ICANN (Internet Corporation for Assigned Names and Numbers)'s contracted Registry Operators are required to act on requests from ICANN (Internet Corporation for Assigned Names and Numbers) within 2 hours of receipt of the request from ICANN (Internet Corporation for Assigned Names and Numbers).



If you believe your name collision meets the criteria above (i.e. your system is suffering demonstrably severe harm as a consequence of name collision or you have a reasonable belief that the name collision presents a clear and present danger to human life), please use the form below to submit your report to ICANN (Internet Corporation for Assigned Names and Numbers).

After submitting the report, please review the **Guide to Name Collision Identification and Mitigation for IT Professionals (https://www.icann.org/en/system/files/files/name-collision-mitigation-01aug14-en.pdf)** [PDF, 476 KB] for more information.

All fields marked with asterisk ("*") are required.

* Domain Name (Domain Name) Causing Harm Related to Name Collision (e.g., foo.bar.example):

* Requestor's Name:

* Requestor's Email:

* Requestor's Phone Number:

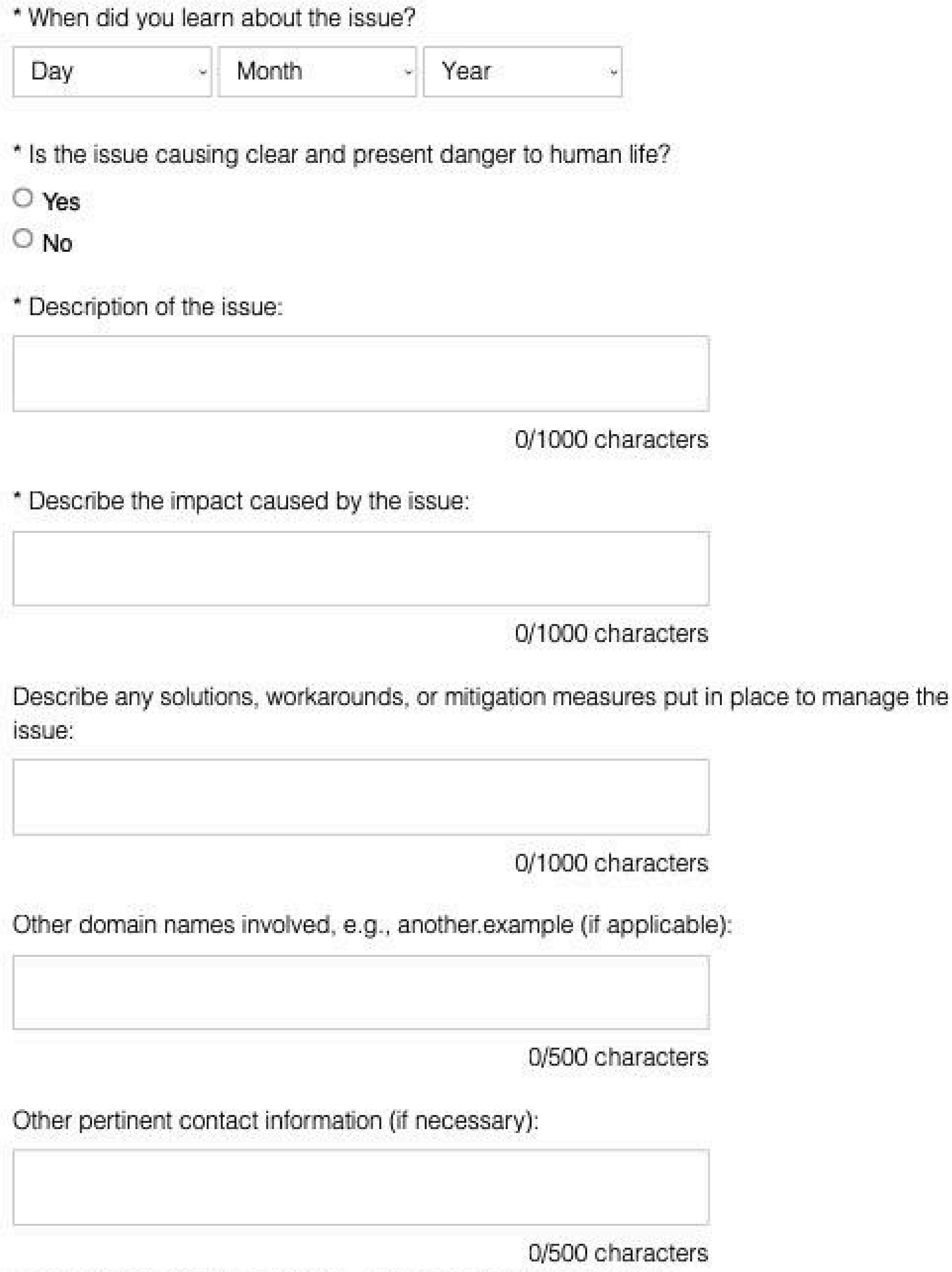
* Requestor's Address:

0/250 characters

Organization Name:

* Start Date of Domain Name (Domain Name) Usage:

Day Month Year



* When did you learn about the issue?

Day ⌄ | Month ⌄ | Year ⌄

* Is the issue causing clear and present danger to human life?

○ Yes

○ No

* Description of the issue:

0/1000 characters

* Describe the impact caused by the issue:

0/1000 characters

Describe any solutions, workarounds, or mitigation measures put in place to manage the issue:

0/1000 characters

Other domain names involved, e.g., another.example (if applicable):

0/500 characters

Other pertinent contact information (if necessary):

0/500 characters

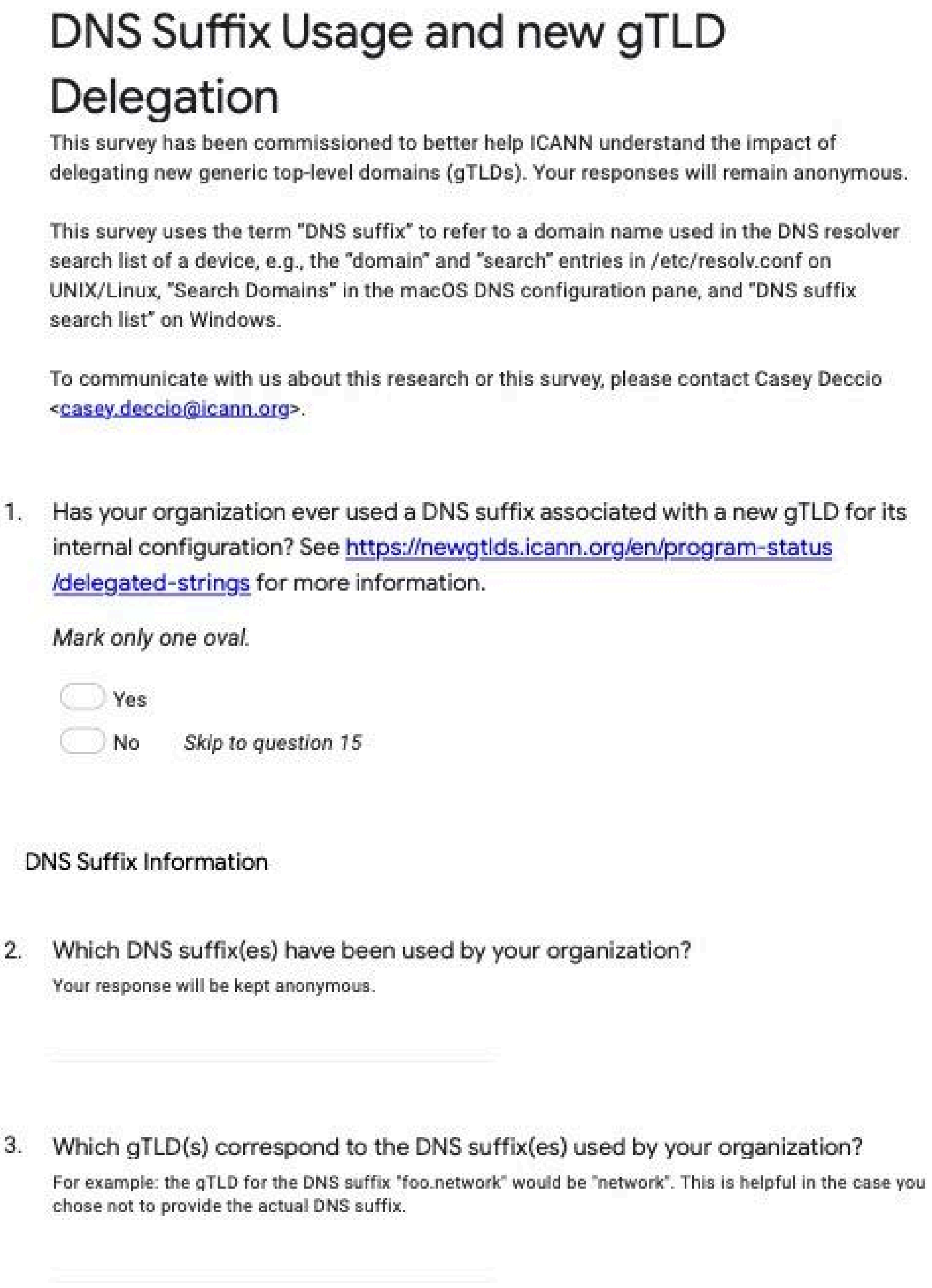
Appendix B - Web Search Results for “127.0.53.53”

Date	Sentiment	gTLD (Delegated)	App	Root Cause Symptoms	ICANN Identified	Other
Sep 2014	Neutral	prod (Aug 2014)	SSH	Unqualified (suffix search list), non-private	Y	
https://serverfault.com/questions/626612/dns-just-started-resolving-my-server-prod-addresses-to-127-0-53-53						
Aug 2015	Neutral	drive (Jun 2015)	Web Browser	Single label resolution	Y	Google search intended
https://superuser.com/questions/958758/why-pinging-drive-gets-replies-from-127-0-53-53						
Oct 2016	Neutral	[Unknown]	[Unknown]	[Unknown]	Y	Firewall logs
https://community.helpsystems.com/forums/intermapper/general-network-questions/3c736b35-b09b-e611-80d8-0050568473e2						
Oct 2014	Neutral	dental (Apr 2014)	[Unknown]	Unqualified (suffix search list, WinXP-style), private	Y	
https://www.reddit.com/r/sysadmin/comments/2jcdso/workstations_resolving_domainlocal_to_12705353/						
Aug 2016	Neutral	dev (Dec 2014)	valet	FQDN, private	Y	Not intended to resolve
https://github.com/laravel/valet/issues/115						
Jan 2016	Neutral	cisco (May 2015)	ping	FQDN and suffix search list, private	N	
https://community.spiceworks.com/topic/1381179-host-name-pinging-to-127-0-53-53						

Feb 2020	Neutral	cpa (Sep 2019)	[Unknown]	FQDN, VPN, private	N	
https://community.meraki.com/t5/Security-SD-WAN/Receiving-127-0-53-53-when-connected-to-the-Client-VPN-FQDN-s/m-p/75929						
Apr 2017	Neutral	[unknown]	RDP	Unqualified (suffix search list), VPN	Maybe (arpa)	
https://community.logmein.com/t5/LogMeIn-Hamachi-Discussions/FQDN-for-hamachi-hosts-127-0-53-53/td-p/139663						
Jun 2015	Neutral	windows (Jun 2015)	[Unknown]	FQDN, private	Y	
https://social.technet.microsoft.com/Forums/en-US/63ac3e27-7e95-47d2-a969-4044737aec0a/dns-collisions-with-windows-tld?forum=winserveripamdhcpdns						
Sep 2014	Angry	prod (Aug 2014)	[Unknown]	Unqualified multi-label, non-private	Y	Google (registry) also known
https://domainincite.com/17278-victims-of-first-confirmed-new-gtld-collision-respond-fuck-google						
Feb 2017	Neutral	bar (Feb 2014)	Apache Kafka (unit testing)	FQDN, private	N	Not intended to resolve
https://issues.apache.org/jira/browse/KAFKA-4765						
Oct 2014	Neutral	[Unknown]	[Unknown]	[Unknown]	Y	
https://blog.51cto.com/u_8378022/1560434						
Aug 2017	Neutral	dev (Dec 2014)	Web browser	FQDN, private	Y	Dev environment
https://apple.stackexchange.com/questions/296588/cant-connect-to-server-app-local-sites						
May 2015	Neutral	int (??) (Nov 1988)	ping	FQDN, ??	Y	
https://blog.manton.im/2015/05/12705353-dns-name-collision.html?m=1						

Oct 2014	Neutral	world (Sep 2014)	php, tnsping	FQDN, private	Y	Access to DB backend; Not intended to resolve
https://crumblybits.com/?p=316						
Dec 2017	Neutral	dev (Dec 2014)	gitlab-ci-multi-runner	FQDN, private	N	
https://gitlab.com/gitlab-org/gitlab-foss/-/issues/41072						
Apr 2017	Neutral	box (Nov 2016)	[Unknown]	FQDN or unqualified, private	Y	Access to pi-hole on LAN
https://discourse.pi-hole.net/t/pi-hole-server-lose-awareness-of-it-self/2715/15						

Appendix C - General Name Collisions Survey



DNS Suffix Usage and new gTLD Delegation

This survey has been commissioned to better help ICANN understand the impact of delegating new generic top-level domains (gTLDs). Your responses will remain anonymous.

This survey uses the term "DNS suffix" to refer to 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.

To communicate with us about this research or this survey, please contact Casey Deccio <casey.deccio@icann.org>.

1. Has your organization ever used a DNS suffix associated with a new gTLD for its internal configuration? See https://newgtlds.icann.org/en/program-status/delegated-strings for more information.

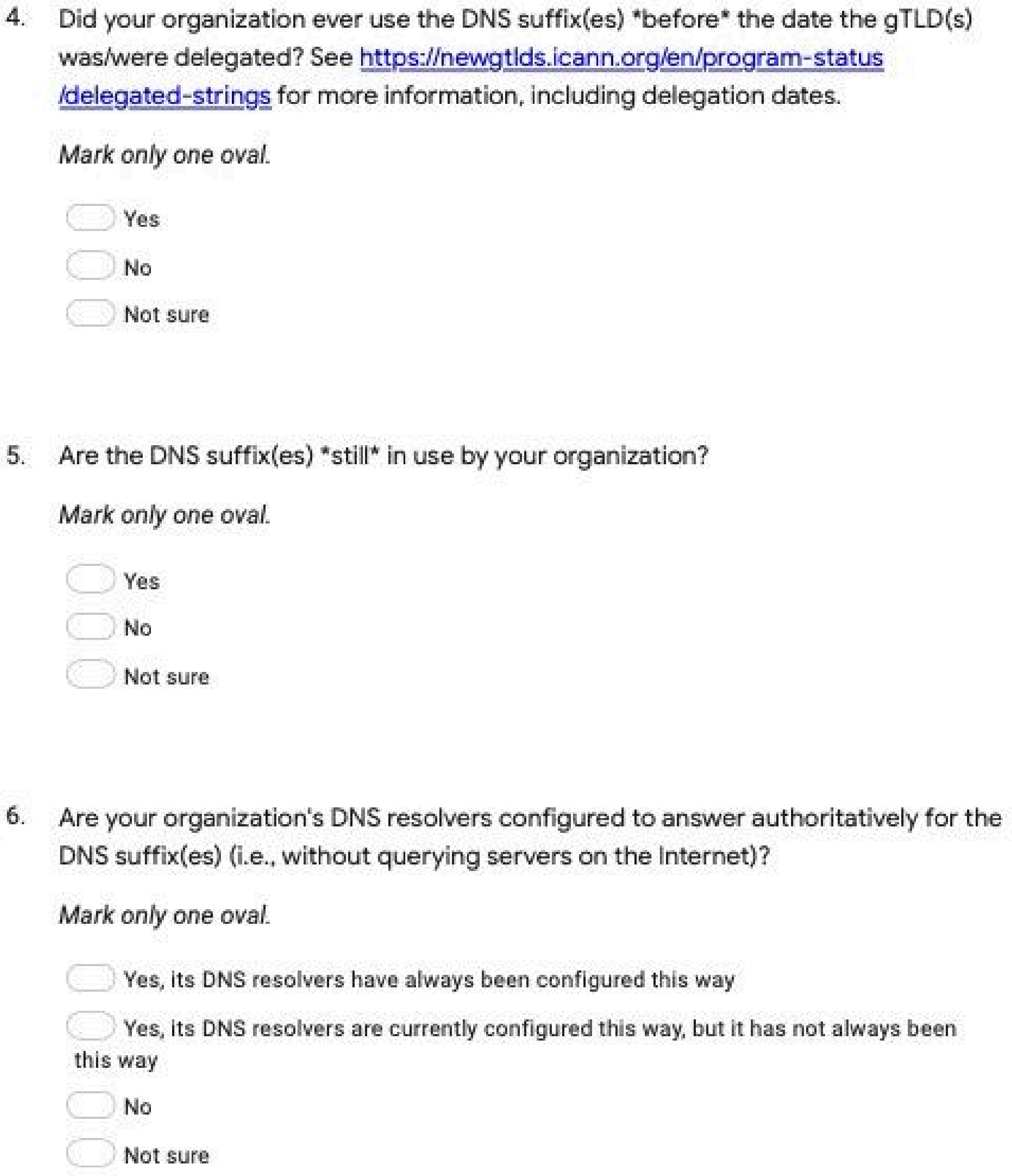
Mark only one oval.

○ Yes
○ No *Skip to question 15*

DNS Suffix Information

2. Which DNS suffix(es) have been used by your organization?
Your response will be kept anonymous.

3. Which gTLD(s) correspond to the DNS suffix(es) used by your organization?
For example: the gTLD for the DNS suffix "foo.network" would be "network". This is helpful in the case you chose not to provide the actual DNS suffix.



4. Did your organization ever use the DNS suffix(es) *before* the date the gTLD(s) was/were delegated? See https://newgtlds.icann.org/en/program-status/delegated-strings for more information, including delegation dates.

Mark only one oval.

○ Yes
○ No
○ Not sure

5. Are the DNS suffix(es) *still* in use by your organization?

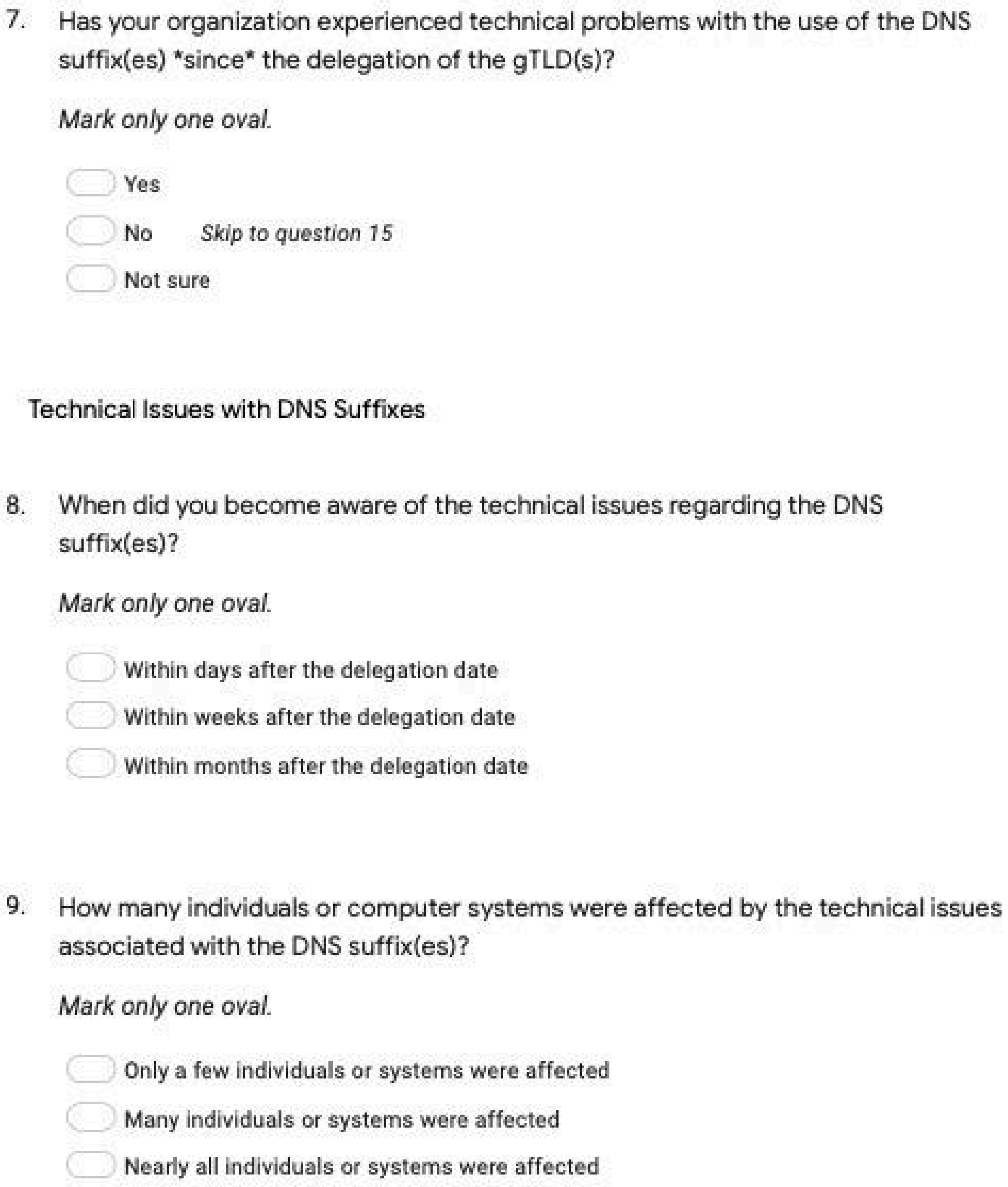
Mark only one oval.

○ Yes
○ No
○ Not sure

6. Are your organization's DNS resolvers configured to answer authoritatively for the DNS suffix(es) (i.e., without querying servers on the Internet)?

Mark only one oval.

○ Yes, its DNS resolvers have always been configured this way
○ Yes, its DNS resolvers are currently configured this way, but it has not always been this way
○ No
○ Not sure



7. Has your organization experienced technical problems with the use of the DNS suffix(es) *since* the delegation of the gTLD(s)?

Mark only one oval.

○ Yes

○ No *Skip to question 15*

○ Not sure

Technical Issues with DNS Suffixes

8. When did you become aware of the technical issues regarding the DNS suffix(es)?

Mark only one oval.

○ Within days after the delegation date

○ Within weeks after the delegation date

○ Within months after the delegation date

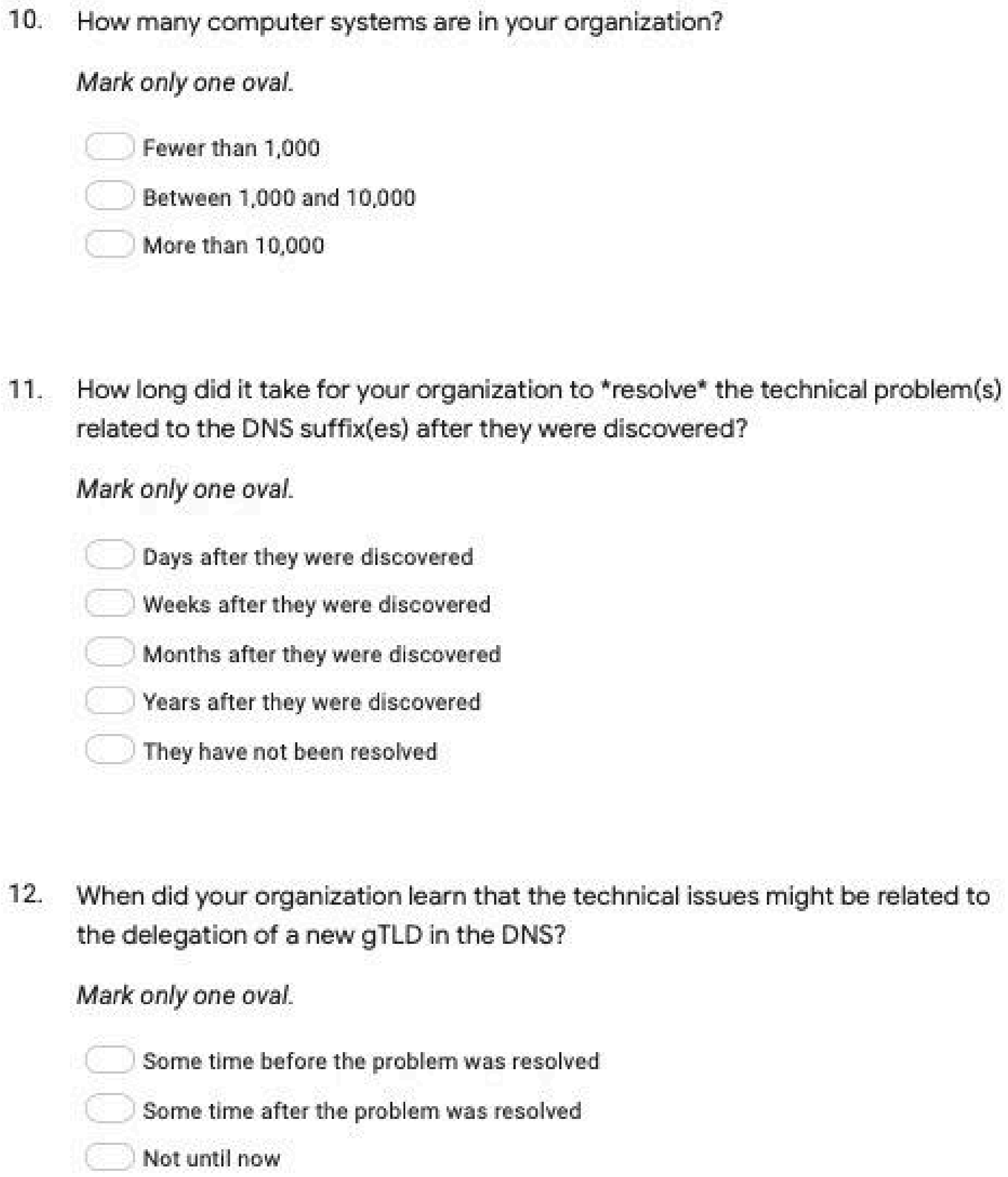
9. How many individuals or computer systems were affected by the technical issues associated with the DNS suffix(es)?

Mark only one oval.

○ Only a few individuals or systems were affected

○ Many individuals or systems were affected

○ Nearly all individuals or systems were affected



10. How many computer systems are in your organization?

Mark only one oval.

○ Fewer than 1,000

○ Between 1,000 and 10,000

○ More than 10,000

11. How long did it take for your organization to *resolve* the technical problem(s) related to the DNS suffix(es) after they were discovered?

Mark only one oval.

○ Days after they were discovered

○ Weeks after they were discovered

○ Months after they were discovered

○ Years after they were discovered

○ They have not been resolved

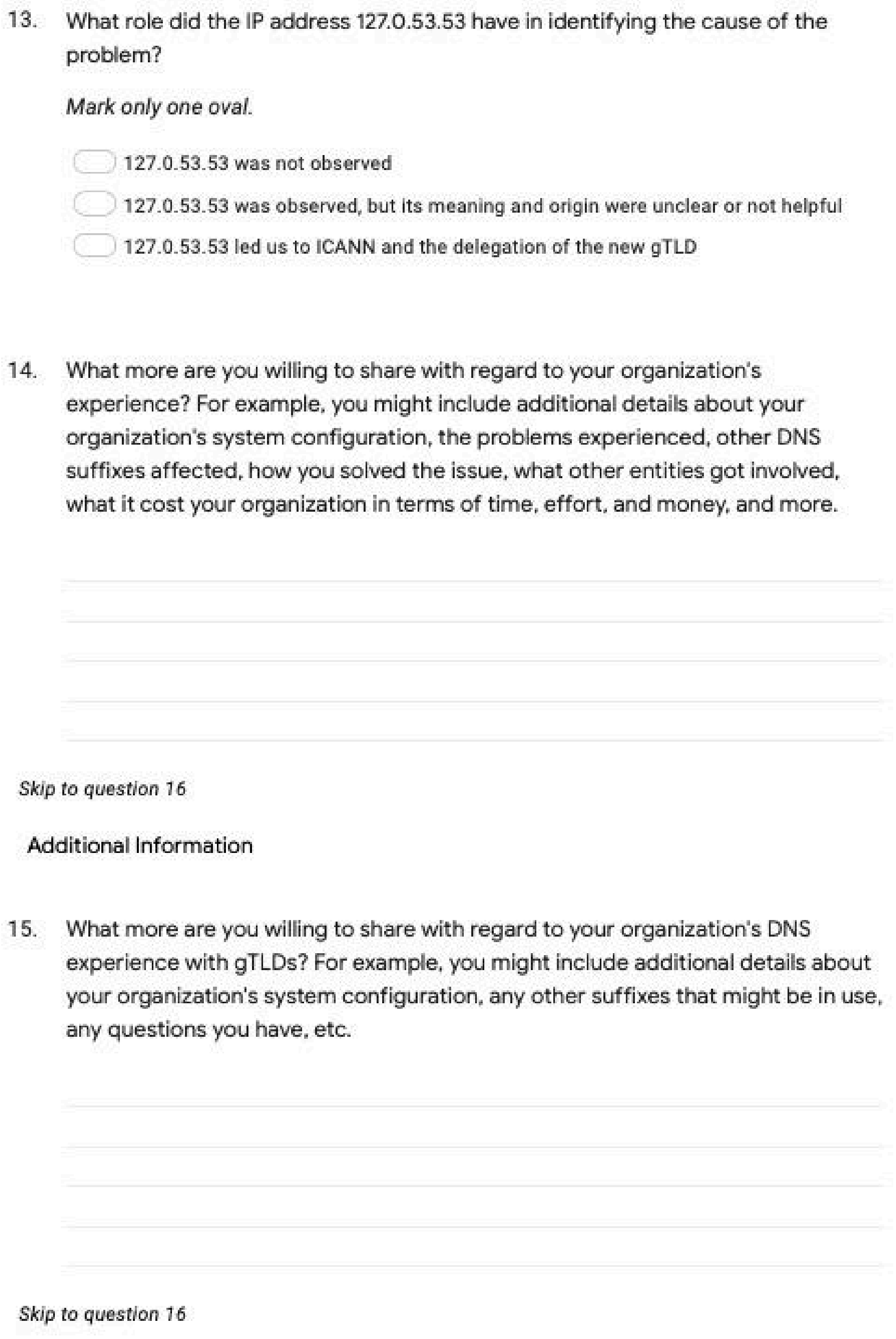
12. When did your organization learn that the technical issues might be related to the delegation of a new gTLD in the DNS?

Mark only one oval.

○ Some time before the problem was resolved

○ Some time after the problem was resolved

○ Not until now



13. What role did the IP address 127.0.53.53 have in identifying the cause of the problem?

Mark only one oval.

○ 127.0.53.53 was not observed

○ 127.0.53.53 was observed, but its meaning and origin were unclear or not helpful

○ 127.0.53.53 led us to ICANN and the delegation of the new gTLD

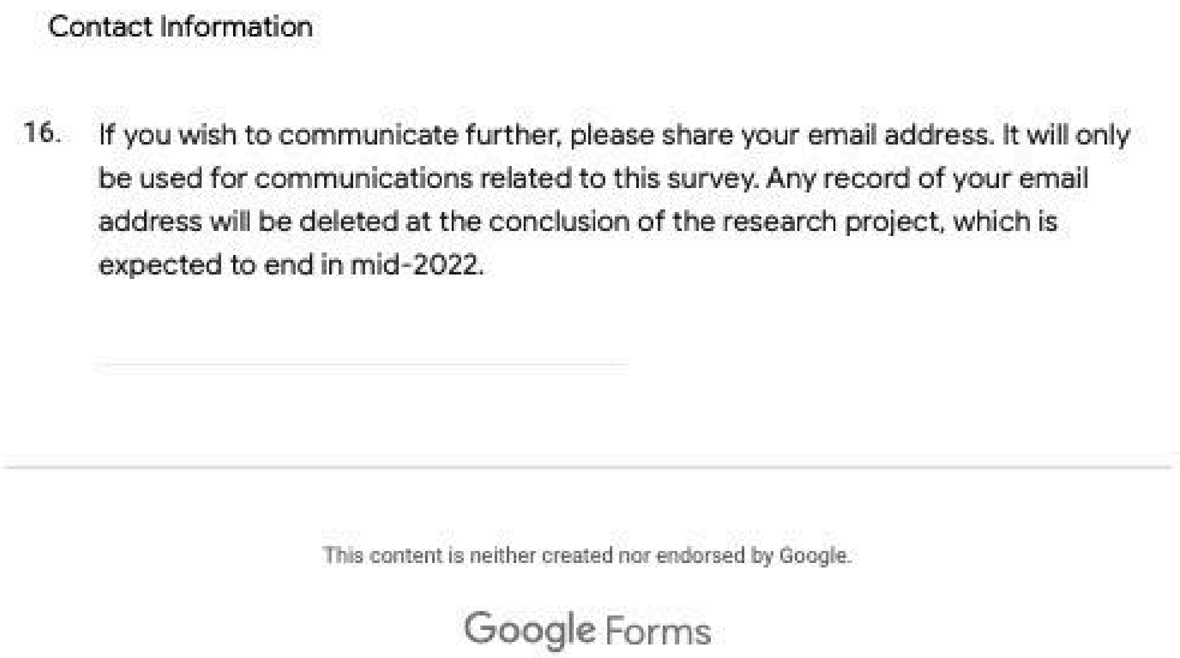
14. What more are you willing to share with regard to your organization's experience? For example, you might include additional details about your organization's system configuration, the problems experienced, other DNS suffixes affected, how you solved the issue, what other entities got involved, what it cost your organization in terms of time, effort, and money, and more.

Skip to question 16

Additional Information

15. What more are you willing to share with regard to your organization's DNS experience with gTLDs? For example, you might include additional details about your organization's system configuration, any other suffixes that might be in use, any questions you have, etc.

Skip to question 16



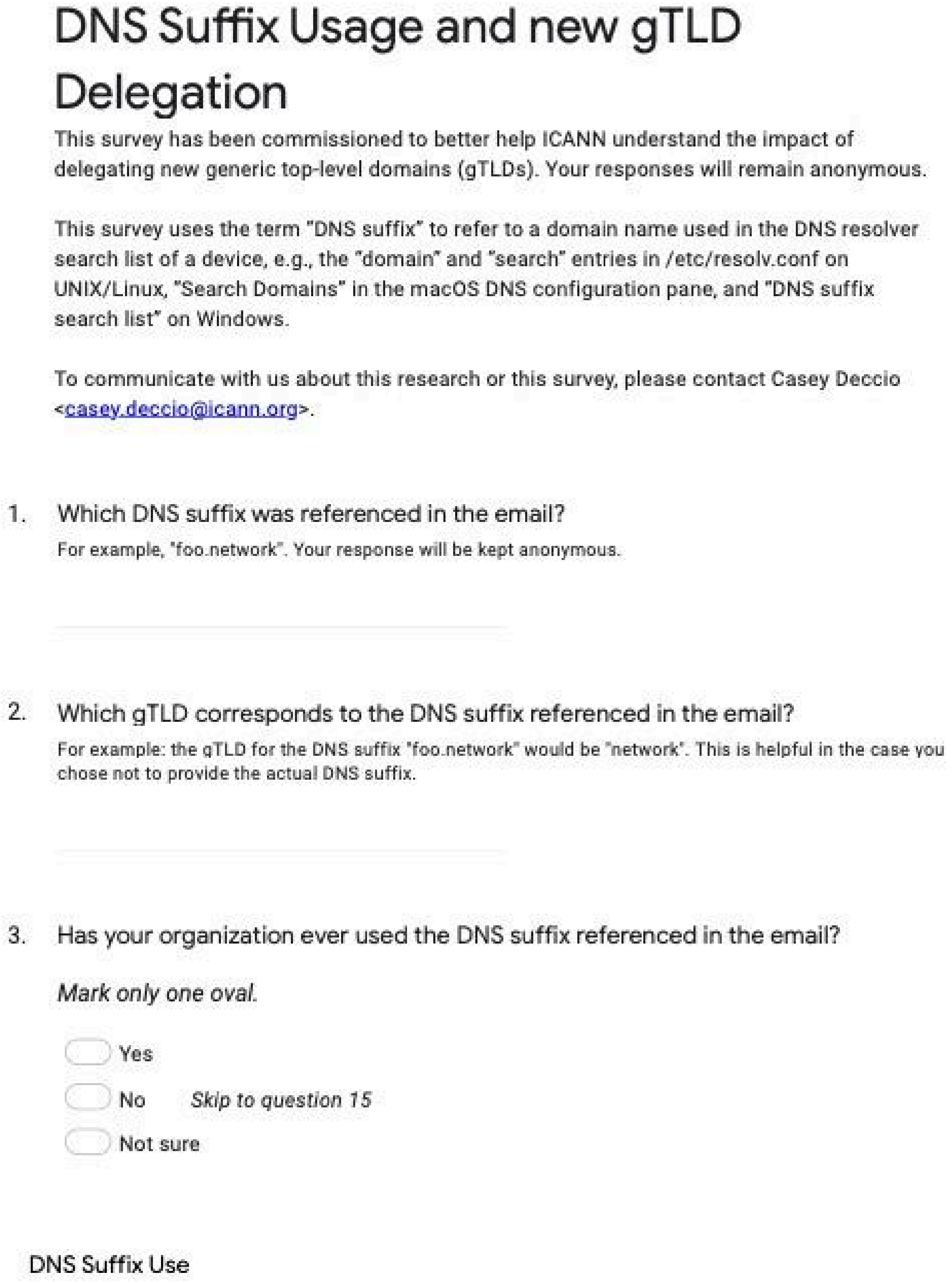
Contact Information

16. If you wish to communicate further, please share your email address. It will only be used for communications related to this survey. Any record of your email address will be deleted at the conclusion of the research project, which is expected to end in mid-2022.

This content is neither created nor endorsed by Google.

Google Forms

Appendix D - Targeted Name Collisions Survey



DNS Suffix Usage and new gTLD Delegation

This survey has been commissioned to better help ICANN understand the impact of delegating new generic top-level domains (gTLDs). Your responses will remain anonymous.

This survey uses the term "DNS suffix" to refer to 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.

To communicate with us about this research or this survey, please contact Casey Deccio <casey.deccio@icann.org>.

1. Which DNS suffix was referenced in the email?

   For example, "foo.network". Your response will be kept anonymous.

2. Which gTLD corresponds to the DNS suffix referenced in the email?

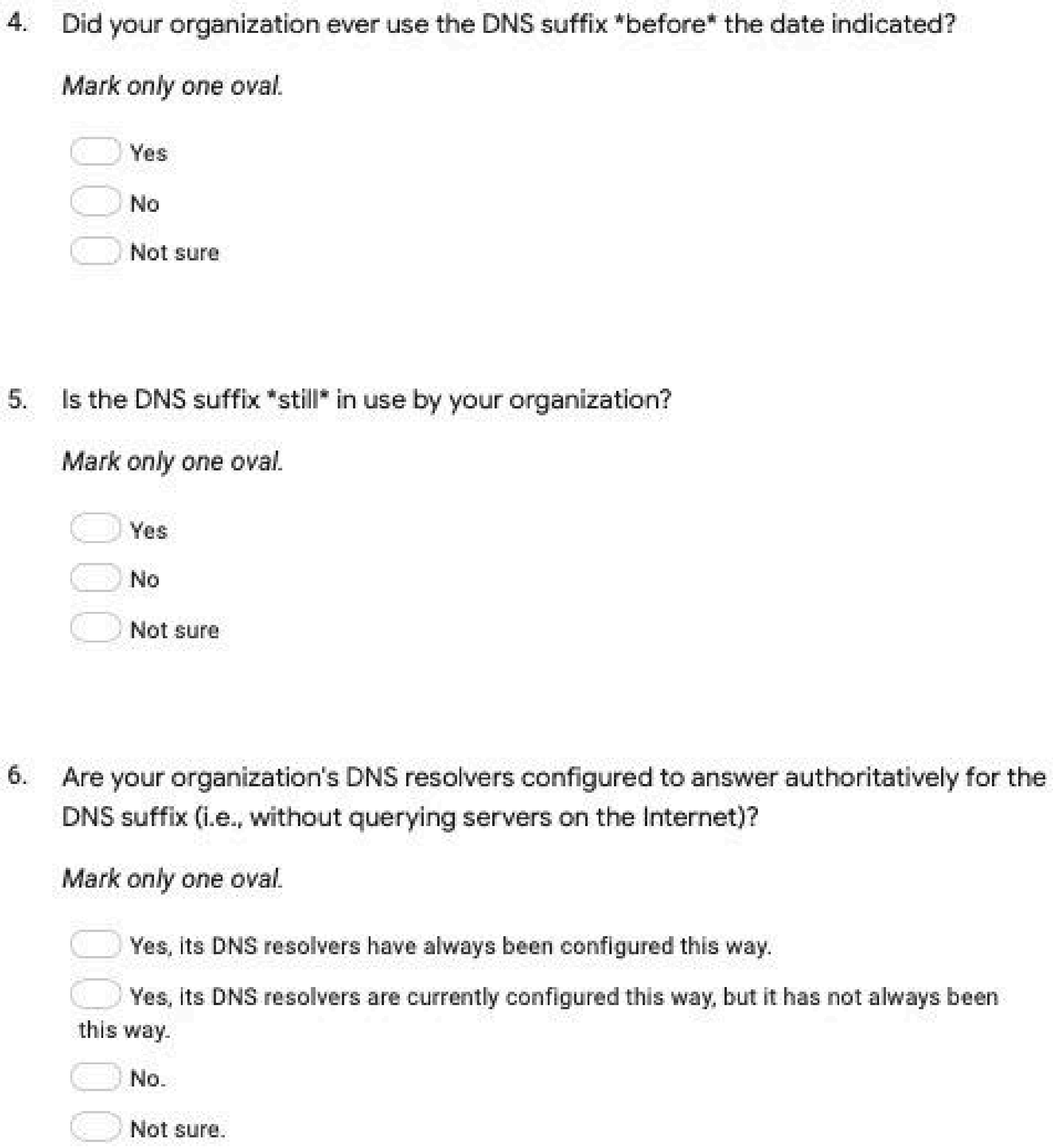
   For example: the gTLD for the DNS suffix "foo.network" would be "network". This is helpful in the case you chose not to provide the actual DNS suffix.

3. Has your organization ever used the DNS suffix referenced in the email?

   Mark only one oval.

   ○ Yes
   ○ No *Skip to question 15*
   ○ Not sure

DNS Suffix Use



4. Did your organization ever use the DNS suffix *before* the date indicated?

Mark only one oval.

○ Yes

○ No

○ Not sure

5. Is the DNS suffix *still* in use by your organization?

Mark only one oval.

○ Yes

○ No

○ Not sure

6. Are your organization's DNS resolvers configured to answer authoritatively for the DNS suffix (i.e., without querying servers on the Internet)?

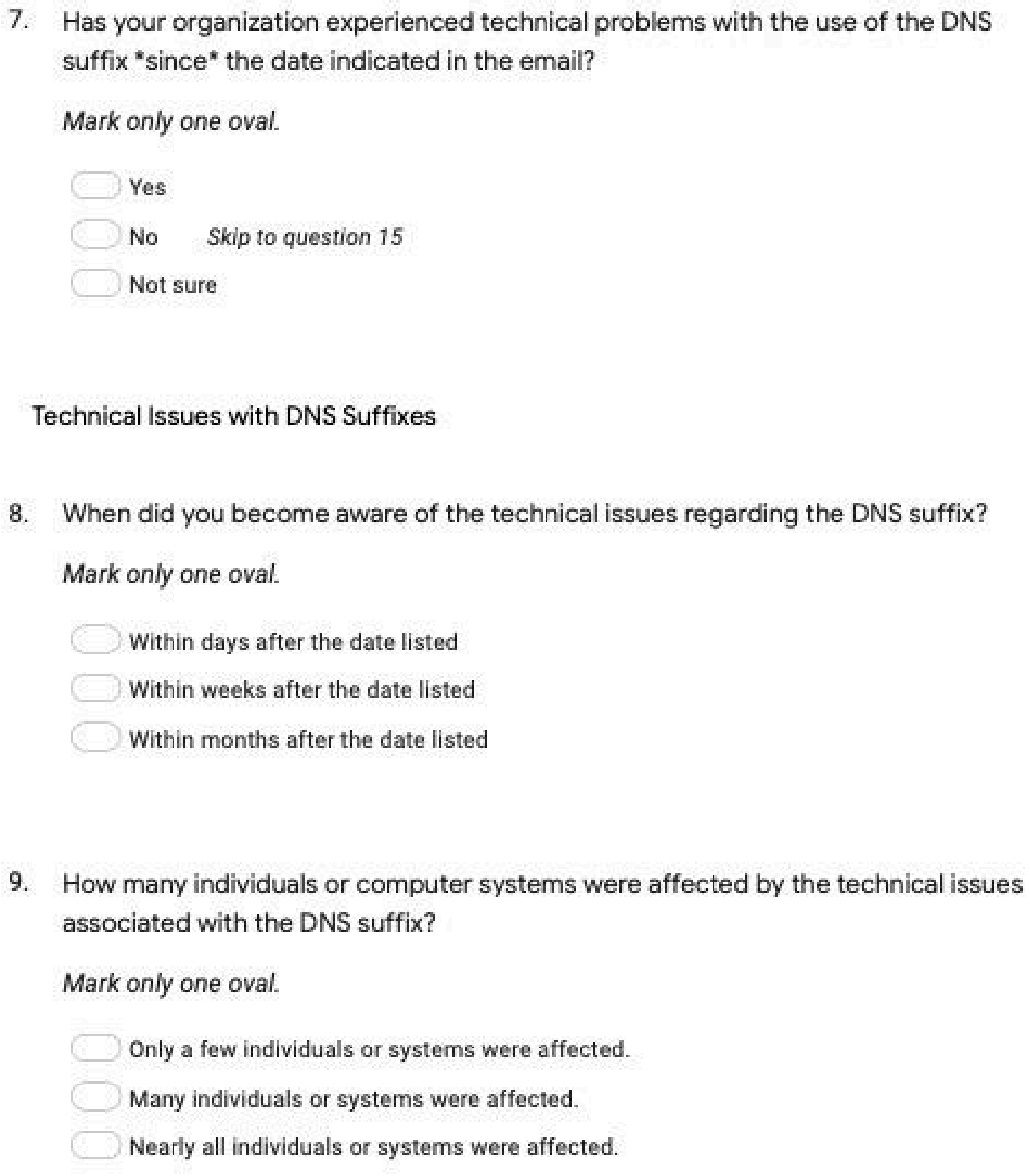
Mark only one oval.

○ Yes, its DNS resolvers have always been configured this way.

○ Yes, its DNS resolvers are currently configured this way, but it has not always been this way.

○ No.

○ Not sure.



7. Has your organization experienced technical problems with the use of the DNS suffix *since* the date indicated in the email?

Mark only one oval.

○ Yes

○ No *Skip to question 15*

○ Not sure

Technical Issues with DNS Suffixes

8. When did you become aware of the technical issues regarding the DNS suffix?

Mark only one oval.

○ Within days after the date listed

○ Within weeks after the date listed

○ Within months after the date listed

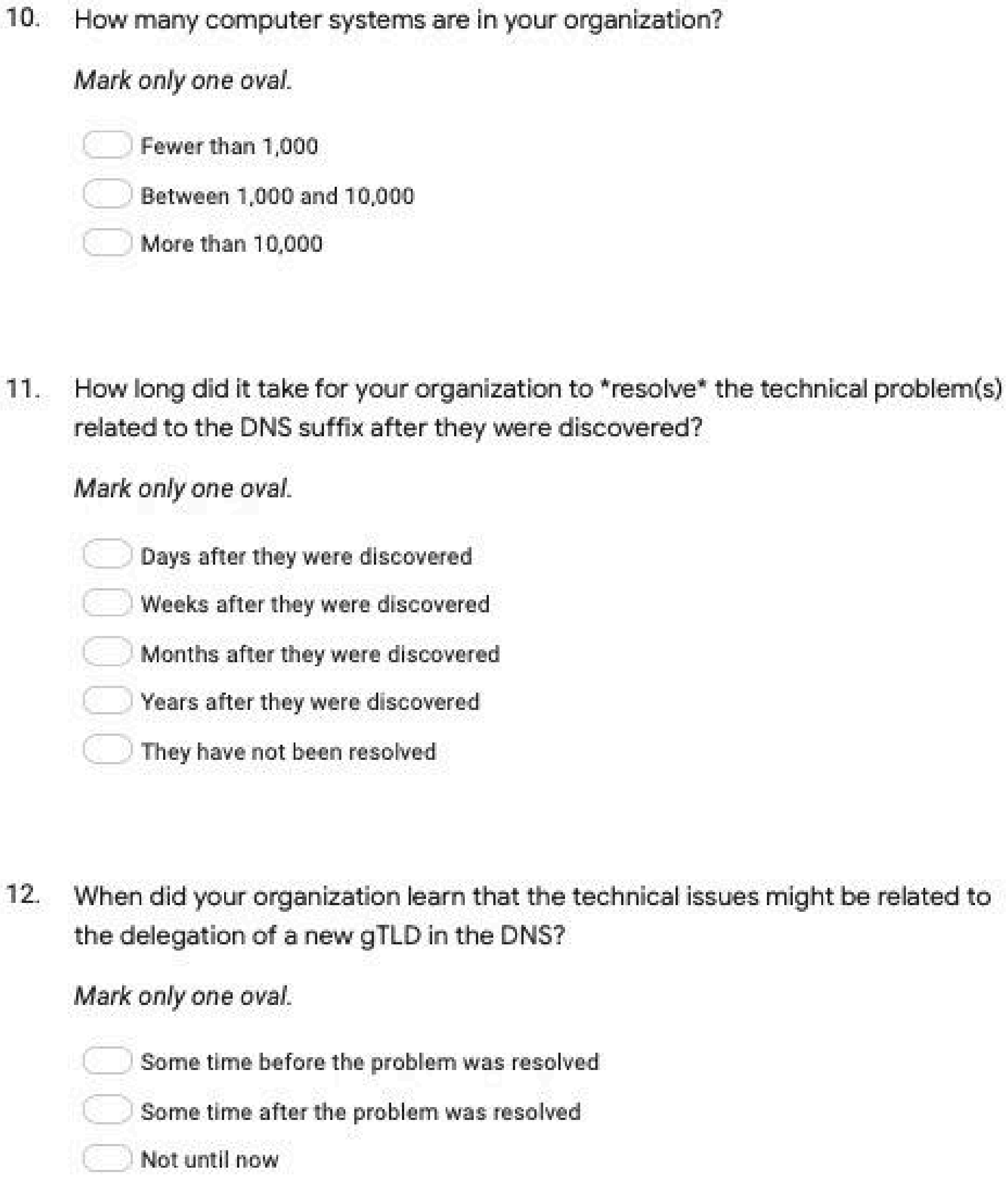
9. How many individuals or computer systems were affected by the technical issues associated with the DNS suffix?

Mark only one oval.

○ Only a few individuals or systems were affected.

○ Many individuals or systems were affected.

○ Nearly all individuals or systems were affected.



10. How many computer systems are in your organization?

Mark only one oval.

○ Fewer than 1,000

○ Between 1,000 and 10,000

○ More than 10,000

11. How long did it take for your organization to *resolve* the technical problem(s) related to the DNS suffix after they were discovered?

Mark only one oval.

○ Days after they were discovered

○ Weeks after they were discovered

○ Months after they were discovered

○ Years after they were discovered

○ They have not been resolved

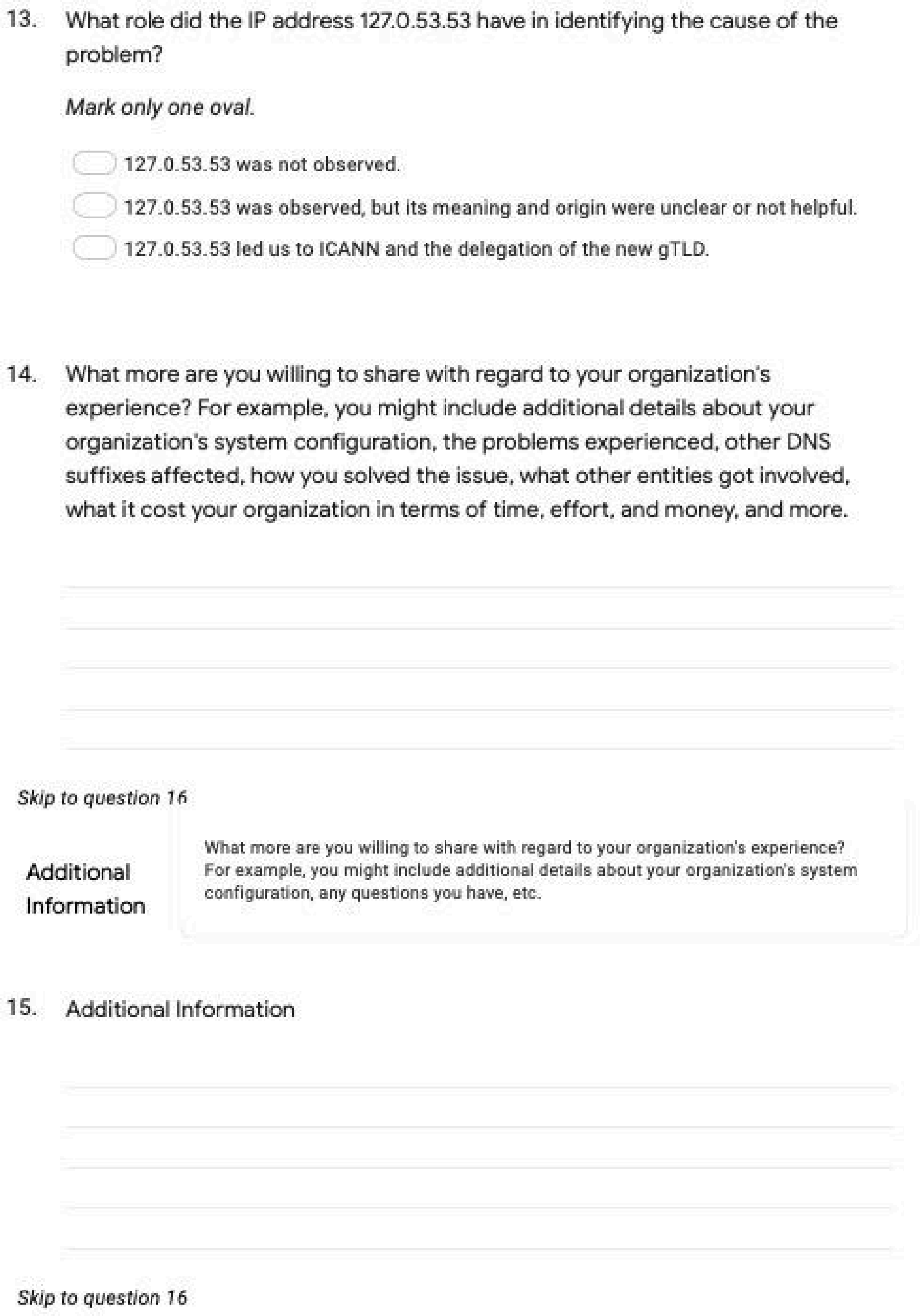
12. When did your organization learn that the technical issues might be related to the delegation of a new gTLD in the DNS?

Mark only one oval.

○ Some time before the problem was resolved

○ Some time after the problem was resolved

○ Not until now



13. What role did the IP address 127.0.53.53 have in identifying the cause of the problem?

Mark only one oval.

○ 127.0.53.53 was not observed.

○ 127.0.53.53 was observed, but its meaning and origin were unclear or not helpful.

○ 127.0.53.53 led us to ICANN and the delegation of the new gTLD.

14. What more are you willing to share with regard to your organization's experience? For example, you might include additional details about your organization's system configuration, the problems experienced, other DNS suffixes affected, how you solved the issue, what other entities got involved, what it cost your organization in terms of time, effort, and money, and more.

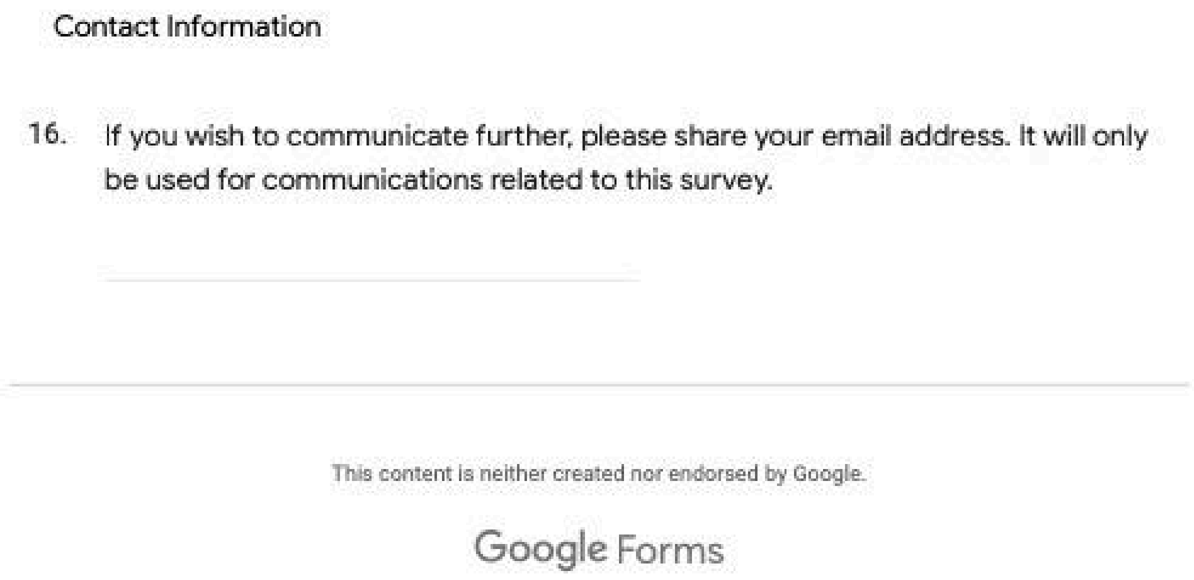
Skip to question 16

Additional Information

What more are you willing to share with regard to your organization's experience? For example, you might include additional details about your organization's system configuration, any questions you have, etc.

15. Additional Information

Skip to question 16



Contact Information

16. If you wish to communicate further, please share your email address. It will only be used for communications related to this survey.

____

This content is neither created nor endorsed by Google.

Appendix E - 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 4b: Root Cause Analysis - New gTLD Collisions

Appendix F - 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

Annex A: Public Comments Analysis

Date Received	Submission	Comment	Changes to Final Report Adopted
2024-02-26	ICANN org	ICANN org welcomes this opportunity to provide input on the NCAP Study 2. The input in the attached document is focused on org's assessment of the implementability of the recommendations should the ICANN Board decide to adopt these, as well as the level of effort and impact of implementation on the new gTLD Program: Next Round. This submission contains two attachments: 1. ICANN org consolidated response to Public Comment on NCAP Study 2.pdf: This document contain's ICANN org's response to this Public Comment and is in lieu of completing the form. https://itp.cdn.icann.org/public-comment/proceeding/Draft%20NCAP%20Study%202%20Report%20and%20Responses%20to%20Questions%20Regarding%20Name%20Collisions-19-01-2024/submissions/ICANN%20org/ICANN%20org%20consolidated%20response%20to%20Public%20Comment%20on%20NCAP%20Study%202-26-02-2024.pdf 2. Visible Interruption (VI) and Visible Interruption and Notification (VIN) - Privacy and data protection review.pdf: This document is referenced in the Public Comment response. https://itp.cdn.icann.org/public-comment/proceeding/Draft%20NCAP%20Study%202%20Report%20and%20Responses%20to%20Questions%20Regarding%20Name%20Collisions-19-01-2024/submissions/ICANN%20org/Visible%20Interruption%20(VI)%20and%20Visible%20Interruption%20and%20Notification%20(VIN)%20-%20Privacy%20and%20data%20protection%20review-26-02-2024.pdf ICANN org fully supports the importance of a mitigation strategy for name collisions. We have concerns about the implementability of some of the proposed recommendations in NCAP Study 2. Furthermore, ICANN org would like to point out that should the ICANN Board decide to direct ICANN org to implement these recommendations, it is likely to have an impact on the resources needed for the next round (compared with those used in the 2012 round to mitigate name collisions) and might have an impact on the implementation timeline of the next round.	Update to agreed-upon definition of "name collision" within the report. Clarification that the NCAP DG does not find it within its remit to provide specific guidance on elements of the operationalization of the Technical Review Team and the Name Collision Risk Assessment Framework, including what data to collect, how to assess this data, and how to maintain compliance with data privacy and risk management standards, with the intention of not prescribing implementation details to provide broad oversight to the ICANN org. Clarification that sufficient resources would be necessary to expeditiously operationalize and implement the Technical Review Team and the Name Collision Risk Assessment Framework, should they be adopted by the ICANN Board, along with the production and distribution of a data privacy and protection policy and appropriate risk mitigation measures for legal compliance.

Date Received	Submission	Comment	Changes to Final Report Adopted
2024-02-27	Ashley Roberts, Com Laude	We thank the NCAP Discussion Group for their work on the Draft NCAP Study 2 and its associated report, and we appreciate this opportunity to comment on that work. We understand that the core recommendation is for a new Name Collision Risk Assessment Framework to replace the Name Collision Management Framework used in the last TLD application round and we appreciate the detailed work on this. We are concerned, however, at the lack of any estimated time frame linked to the Name Collision Risk Assessment Framework workflow. A previous iteration of the workflow, presented last Autumn, did include estimated timings for the different stages of the workflow, as did the Controlled Interruption process used during the last round of new gTLDs. While we understand it may not be possible to be absolute in predicting a time frame for conducting the risk assessment for a TLD, it is important that applicants have some idea of how long a “typical” risk assessment is likely to take, with the understanding that if issues are discovered then the assessment may take longer. Therefore, we would urge you to consider attaching estimated time frames to the risk assessment workflow to provide a level of predictability for applicants. In addition, we suggest the risk assessment process should begin as soon as possible following the publication of the applied-for TLDs in the next round, running in parallel with the application evaluation and prior to other associated processes (e.g. objections, contention resolution, etc.). This will help the efficiency of the overall process, helping to avoid applications being rejected on the basis of name collision issues after the applicant has already spent considerable time and resources on navigating other obstacles, such as objections. Further, whether there is a name collision risk requiring mitigation is likely to be a factor applicants would take into consideration when seeking to resolve contention. Thank you for your consideration of these comments. We urge the Discussion Group to attach estimated time frames to the name collision risk assessment workflow outlined in the report to provide some predictability for applicants. We also ask you to consider advising that the risk assessment analysis commence as soon as possible after the publication of the applied-for TLDs, and certainly prior to other TLD assessment procedures such as objections and contention resolution.	Clarification that, should they be adopted by the ICANN Board, the operationalization of the Technical Review Team and the implementation of the Name Collision Risk Assessment Framework be completed expeditiously, for which sufficient resources must be provided. Specify that time frames for the Name Collision Risk Assessment Framework be distributed to the public as early as possible, should the recommendation be adopted.

Date Received	Submission	Comment	Changes to Final Report Adopted
2024-02-28	ICANN Business Constituency (BC)	Our attachment is in lieu of completing this form. The BC thanks the NCAP DG for their significant efforts to assess and detail the challenges posed by Name Collision (NC) and possible solutions. While we agree with most of the assessment, we have some concerns and suggestions regarding the new Risk Assessment Framework. Please see our attached comment. https://itp.cdn.icann.org/public-comment/proceeding/Draft%20NCAP%20Study%202%20Report%20and%20Responses%20to%20Questions%20Regarding%20Name%20Collisions-19-01-2024/submissions/ICANN%20Business%20Constituency%20(BC)/BC%20Comment%20on%20Name%20Collision%20Study%202-28-02-2024.pdf	Clarification that, should they be adopted by the ICANN Board, the operationalization of the Technical Review Team and the implementation of the Name Collision Risk Assessment Framework be completed expeditiously, for which sufficient resources must be provided. Specification that all strings be subject to a typical technical evaluation process without preferential review treatment for any grouping of strings. The implementation of special procedures for certain types of strings based upon policy adoption is out of scope for this report. Clarification that data that is presently available to the public, which applicants could use to self-assess their applications is constrained.
2024-02-28	Rubens Kuhl	The dichotomy suggested in the report between IPv6 and Controlled Interruption is not based on fact-based finding, but on lack of testing. ::1 (meaning ::1/128 as in IPv6 there is no localhost subnet, only a localhost) is a perfectly good solution to add IPv6 support to Controlled Interruption, targeting IPv6-only hosts. I support doing a study with a few key operating systems to confirm its usefulness and lack of side effects before the final report is published, so it gets quicker to a name collision framework without reconvening NCAP DG. On VI/VIN, the staff analysis of privacy risks makes a strong case for not adopting VIN at all. But VI seems possible, so some work on VI (notably on defining possible legal basis) can increase the odds of VI making part of the final framework. While I'm personally not a strong supporter of VI, the decision of including VI or not should be based on its merits, and there seems to be a number of DG members that believe it has merits.	Clarification that the proposed data collection methods were deliberated upon by the NCAP DG based upon data privacy risks and potential benefits and that the data collection methods proposed for the TRT are a small sampling of known and tested methods. Other methods may be used, but they remain untested and are out of scope within this report. Ultimately, which methods to use should be critically considered during the operationalization of the TRT. Clarification that IPv6 is a risk tradeoff that was thoroughly discussed in the JAS report, and that there is no clear, risk-free approach to 2012-style CI in v6 space.

Date Received	Submission	Comment	Changes to Final Report Adopted
2024-02-28	Intellectual Property Constituency	These comments are submitted by the Intellectual Property Constituency (“IPC”), whose membership includes and represents trade associations, large multinational corporations, as well as small businesses and individuals. The IPC appreciates the opportunity to submit the following comments in connection with the Draft Name Collision Analysis Project (NCAP) Study 2 Report and the proposed responses to ICANN Board questions. We note that the IPC's comments do not encompass the technical findings stated in the report. Our comments are limited to a general overview of certain issues the membership believes should be addressed by the NCAP Discussion Group prior to referring the report to the Security and Stability Advisory Committee (SSAC). The IPC notes that the NCAP Study 2 Report does not make a recommendation as to whether a string that is designated as a Collision String by the Technical Review Team's assessment after test delegation to the root (and before contract award) should be removed from the root. Specifically, this would be a removal after the initial delegation for risk purposes. The IPC encourages the NCAP Discussion Group to specify a recommendation in this regard, even if that recommendation is simply that the Technical Review Team should make the determination whether to leave the string in the root or to remove it. The IPC also notes that there is no specific recommendation in the Study 2 Report as to whether the Name Collision Risk Assessment Framework should be applied to a particular string before or after the Resolution of other evaluations and other ICANN processes such as Objections and/or String Contention. Given that name collision issues may be an important part of the assessment by an applicant as to whether to move forward with any given application, the IPC recommends that the NCAP Discussion Group modify the Study 2 Report to specify that the Name Collision Risk Assessment be conducted as soon as possible after it is determined that the applicant meets other technical and financial requirements. In this manner, expensive Objection and String Contention proceedings may be either avoided or resolved at an early stage in the process of bringing the TLD forward to contract award. The IPC further suggests that the NCAP Discussion Group consider a potential situation if a .brand TLD is found to collide with its own internal TLD. In such instances, there should be accommodation for that TLD operator to implement the mitigation measures that it deems necessary to alleviate any effects of such collision, if any. Lastly, with respect to the Recommendation not to proceed to conduct Study 3 in relation to mitigation efforts, the IPC supports that Recommendation. We understand that considerations of name collision risk occurring in the interaction between the DNS and various alternate root environments as described in SAC 123 is out-of-scope for the current NCAP work. These collisions nevertheless remain a matter of concern in the long term. The NCAP Discussion Group may wish to consider whether it is appropriate to recommend further study on this topic. In the view of the IPC, any such further study should not delay the timing for the next new gTLD application round. IPC provides comments on the non technical aspects of the NCAP Study 2 report. The comments are intended to encourage more definitive actions including the timing of the collision review in relation to other ICANN reviews and processes, determining whether to leave a string in the root or remove it and to expedite review to in order to avoid costly mitigation in the future. Further, any additional studies should not delay the timing of the next round.	Clarification that, should they be adopted by the ICANN Board, the operationalization of the Technical Review Team and the implementation of the Name Collision Risk Assessment Framework be completed expeditiously, for which sufficient resources must be provided. Clarification that, there must be a process for–after test delegation to the root zone–the removal of a string from the root upon its addition to the Collision String List following review by the TRT. Specification that all strings be subject to a typical technical evaluation process without preferential review treatment for any grouping of strings. The implementation of special procedures for certain types of strings based upon policy adoption is out of scope for this report.

Date Received	Submission	Comment	Changes to Final Report Adopted
2024-02-28	Registries Stakeholder Group (RySG)	The Registries Stakeholder Group (RySG) welcomes the opportunity to comment on the Draft NCAP Study 2 Report and Responses. We appreciate the time and expertise the participants dedicated to developing responses to questions regarding name collisions. We also encourage the Board to take into account ICANN staff's contribution and analysis of privacy issues. As evidenced by our engagement in community efforts, the RySG has experience and interest in addressing privacy concerns. We appreciate the Board's consideration of the issue and are happy to share our expertise and experience as appropriate. The RySG supports the Board maintaining momentum on these recommendations while looking for constructive and efficient ways to continue to examine the highlighted concerns. As this topic has been identified as on the critical path for the ongoing Subsequent Procedures work, we encourage prompt review by the Board.	None.
2024-02-28	ALAC Policy staff in support of the At-Large Community	Please find attached (PDF) the ALAC Statement on Draft NCAP Study 2 Report and Responses to Questions Regarding Name Collisions. Ratification information is included on the cover page. Kind Regards, ICANN Policy Staff in support of the At-Large Community https://itp.cdn.icann.org/public-comment/proceeding/Draft%20NCAP%20Study%202%20Report%20and%20Responses%20to%20Questions%20Regarding%20Name%20Collisions-19-01-2024/submissions/policy%20staff%20in%20support%20of%20the%20at-large%20community-at-large%20advisory%20committee%20(alac)/AL-ALAC-ST-0124-01-00-EN-28-02-2024.pdf Summary: The ALAC supports the recommendations and findings provided in the Name Collision Analysis Project (NCAP) Discussion Group's study 2 report and the detailed responses to the Board's questions regarding name collisions. The ALAC agrees that NCAP study 3 should not proceed at this time. Furthermore, the ALAC agrees with the overarching assertion that name collision is a risk management issue and supports the NCAP DG's call for an independent and neutral Technical Review Team. Additionally, the ALAC agrees that the best available data should be available to the Technical Review Team when strings are being assessed. The ALAC notes that though the completion of NCAP Study 2 should no longer be an impediment to the opening of the Next Round Program, there are significant recommendations stemming from this study, such as the establishment of a Technical Review Team and the development and documentation of an emergency change process, that, if adopted by the ICANN Board, must be implemented expeditiously as to not delay the next round of new gTLDs.	Clarification that, should they be adopted by the ICANN Board, the operationalization of the Technical Review Team and the implementation of the Name Collision Risk Assessment Framework be completed expeditiously, for which sufficient resources must be provided.