Name Collision Analysis Project Study Two Report DRAFT

19 January 2024

Table of Contents

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.[2] 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

[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

[2] See RFC 8198: Aggressive Use of DNSSEC-Validated Cache, https://www.rfc-editor.org/rfc/rfc8198

collision. Furthermore, the DG also evaluated gaps in the availability and completeness of data 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 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.

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.[3] 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.[4] 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.[5]

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:

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

[3] 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 n-board-02-11-2017-en#2.a.rationale

[4] 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

[5] 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 n-board-25-03-2021-en#2.b

And the final tasks included:

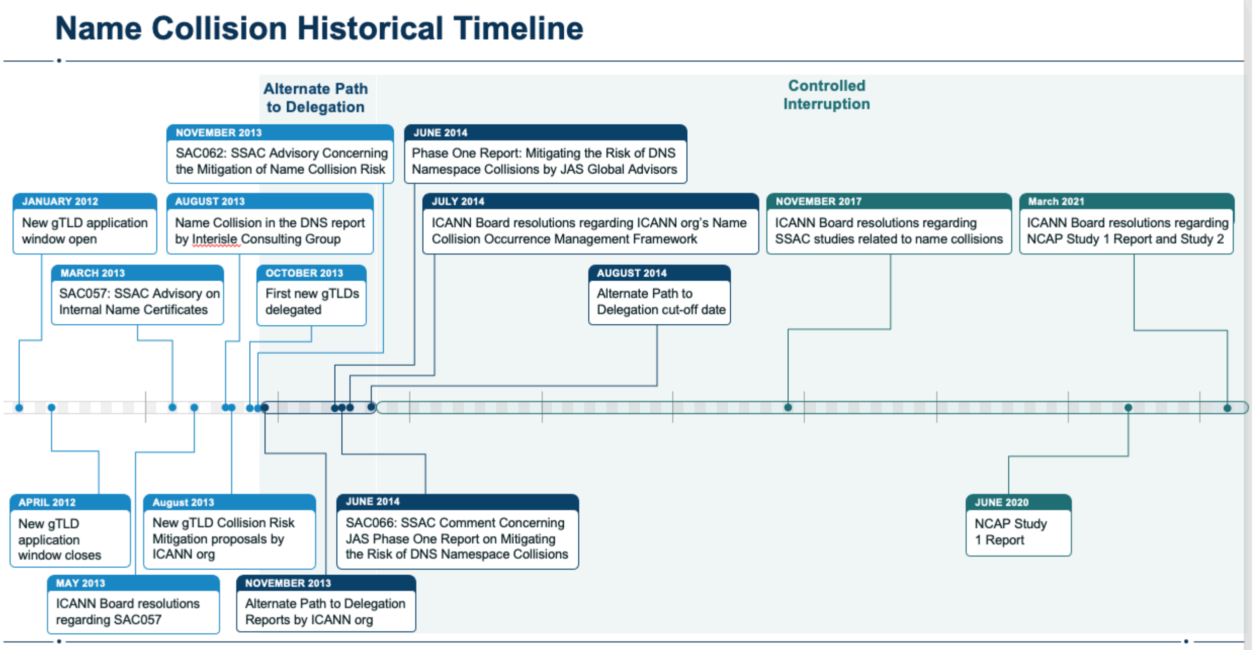
Task	**Steps**	**Responsible Party**
Study of ICANN Collision Reports	Perform an analysis of ICANN Collision Reports to determine the underlying cause of these collisions.	Technical Investigator
	Produce a report on the results of the analysis.	
Impact and Data Sensitivity Analyses	Research the impact 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.[6] 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.[7]

> A name collision occurs when an attempt to resolve a name used in a private name space[8] (e.g. under a non-delegated Top-Level Domain, or a short, unqualified name) results in a

[6] See ICANN Wiki: Name Collision, https://icannwiki.org/Name_Collision
[7] See ICANN, Name Collision Resources & Information, https://www.icann.org/resources/pages/name-collision-2013-12-06-en
[8] 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.[9] 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.[10] 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.[11]

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.[12] 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.

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

[10] 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

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

[12] 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.[13] 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.

[13] 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.[14]

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 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.

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

- 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.”[15] The Board approved this proposal for implementation and outreach via resolutions 2013.10.07.NG01 - 2013.10.07.NG02.[16] 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[17], 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 a trial delegation of some form. Details of the proposed methods can be found in Section 3.2 of the proposal.

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

[16] 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

[17] 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.”

- 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".[18] 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.[19]

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,"[20] a commissioned report by JAS Global Advisors (hereinafter described as the "JAS Report"); the final report was published in 2015.[21] ICANN used the JAS Report, which primarily relied upon DITL data analysis, to develop the "Name Collision Occurrence Management Framework[22]," a guide for ICANN and the new gTLD registry operators on how to handle name collisions. The report includes several recommendations immediately relevant to the Study Two report; we refer the reader to the JAS Report for the supporting analysis associated with each recommendation.

[18] 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

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

[20] 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

[21] 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

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

- **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."[23] 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.

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

[23] 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

- 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,"[24] originally published on 30 July 2014 and approved and directed for implementation by the ICANN Board with Resolution 2014.07.30.NG01[25]. 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 "will work within the IETF and with other relevant technical communities to identify a

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

[25]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

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.[26] 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.[27] The final SSAC NCAP Proposal, published in September 2018, was organized into three studies.[28] 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.

[26] 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

[27] 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

[28] 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[29] on 14 March 2019 with Resolutions 2019.03.14.20 – 2019.03.14.23.[30]

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.[31] 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.

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

[30] 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

[31] 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[32], 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[33] 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.

[32] 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.”

[33] 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[34]. On 5 February 2021,

[34] 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[35] 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.[36] 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.[37] 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.

[35] 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

[36] 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

[37] 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[38], ICANN75[39], ICANN76[40], ICANN77[41] and ICANN78[42] to ensure the broader community was aware of the work, findings, and pending recommendations.

[38] See ICANN74: NCAP Status Update, https://74.schedule.icann.org/meetings/wcin8eB2MQNNRwWP6
[39] See ICANN75: NCAP Final Update: Preparation for Public Comment, https://75.schedule.icann.org/meetings/WxsCLa9h4NapEaq6n
[40] See ICANN76: Name Collision Analysis Project (NCAP): Study 2 Update, https://icann76.sched.com/event/1IfwG/name-collision-analysis-project-ncap-study-2-update
[41] See ICANN77: Name Collision Analysis Project Study 2 Update, https://icann77.sched.com/event/1N5ZJ/name-collision-analysis-project-study-2-update
[42] See ICANN78: Name Collision Analysis Project Study 2 Update, https://icann78.sched.com/event/1Sgpj/name-collision-analysis-project-study-2-update

Mention something about Sec. 3.6 (privacy considerations, group consensus)

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 Strings - a string that manifests name collisions
  - Collision String List - a list of names not to be allocated nor delegated (on the collision string list).[43]
- 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."[44]
- 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[45] 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[46]. 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

[43] 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

[44] 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

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

[46] 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

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
- 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." [47]
- 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.[48]

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:

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

[48] 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

> 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 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.[49]

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.

[49] 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[50]
- A Perspective Study of DNS Queries for Non-Existent Top-Level Domains[51]
- Root Cause Analysis: wpad.domain.name[52] and New gTLD Collisions[53]

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.

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

[51] 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

[52] See Root Cause Analysis - wpad.domain.name, https://www.icann.org/en/system/files/files/root-cause-analysis-wpad-18jan23-en.pdf

[53] See Root Cause Analysis - New gTLD Collisions, https://www.icann.org/en/system/files/files/root-cause-analysis-new-gtld-collisions-18jan23-en.pdf

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 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.[54][55]

[54] See Root Cause Analysis - New gTLD Collisions, https://www.icann.org/en/system/files/files/root-cause-analysis-new-gtld-collisions-18jan23-en.pdf

[55] See Root Cause Analysis - wpad.domain.name, https://www.icann.org/en/system/files/files/root-cause-analysis-wpad-18jan23-en.pdf

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[56] and the Board resolutions 2021.03.25.11 – 2021.03.25.14[57] 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[58].

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) Data Sets: 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.

[56] 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

[57] 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

[58] 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.[59]

On 4 August 2021, Alain Durand from the Office of the Chief Technology Officer at ICANN and Christian Huitema from Private Octopus presented[60] 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[61], 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[62]), specifically as part of the ICANN DNS Magnitude project.[63] 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.

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

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

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

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

[62] 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

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

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.

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.[64] 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.”

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 name collision string

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

Along with these four (4) major classes of measurement, other characteristics identified as Critical Diagnostic Measurements include[65]:

- 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.

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

[65] 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.

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."[66]

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 given class, but [there] are no records of the given type"[67] (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.[68] With NODATA responses, resolvers are forced to

[66] 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
[67] See RFC 2308: Negative Caching of DNS Queries (DNS NCACHE), https://www.rfc-editor.org/info/rfc2308
[68] 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.

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.

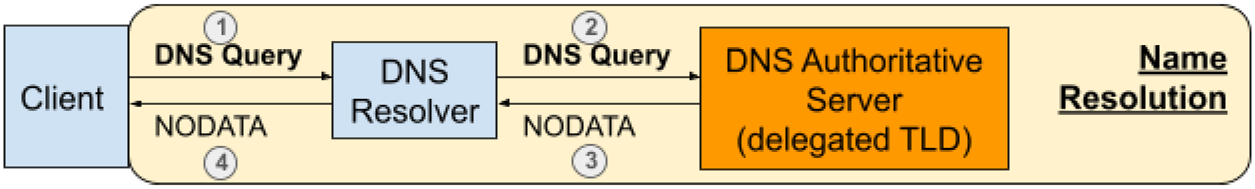


Figure 2: Representation of DNS NODATA response (“No Interruption”)

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[69]. 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[70]. 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.

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

[69] 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
[70] See RFC 2308: Negative Caching of DNS Queries (DNS NCACHE), https://www.rfc-editor.org/info/rfc2308

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.

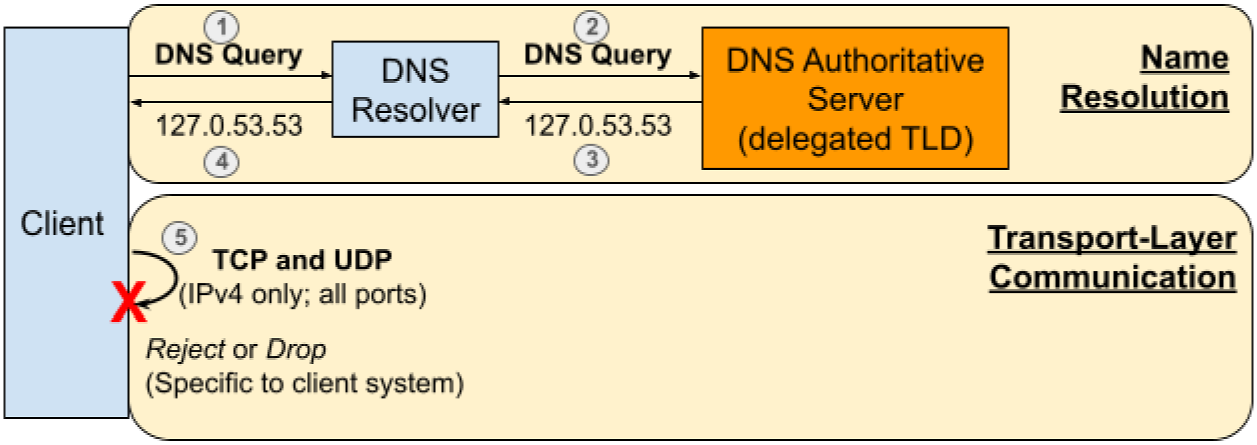


Figure 3: Representation of Transport-Layer Rejection at Local System ("Controlled Interruption")

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. 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").

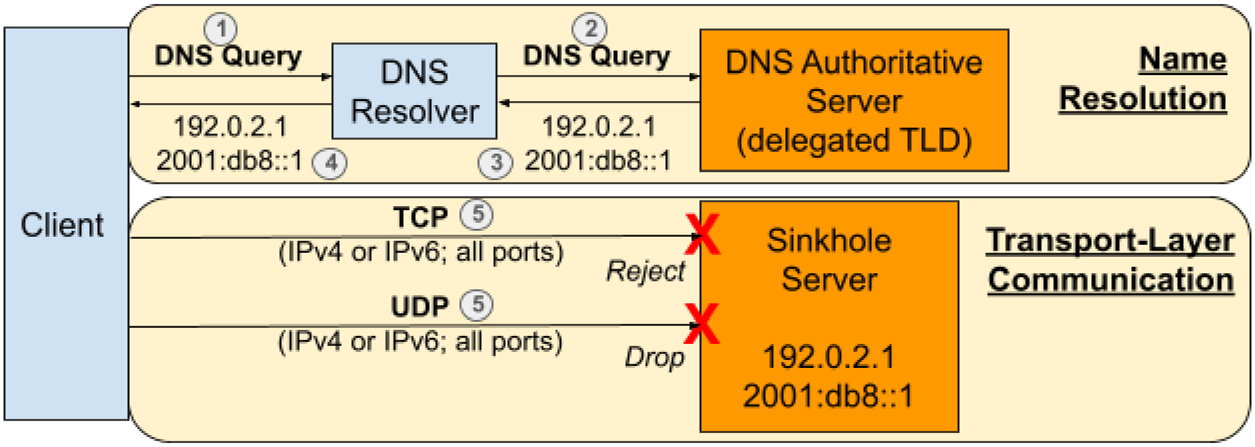


Figure 4: Representation of Transport-Layer Rejection at Public IP ("Visible Interruption")

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").

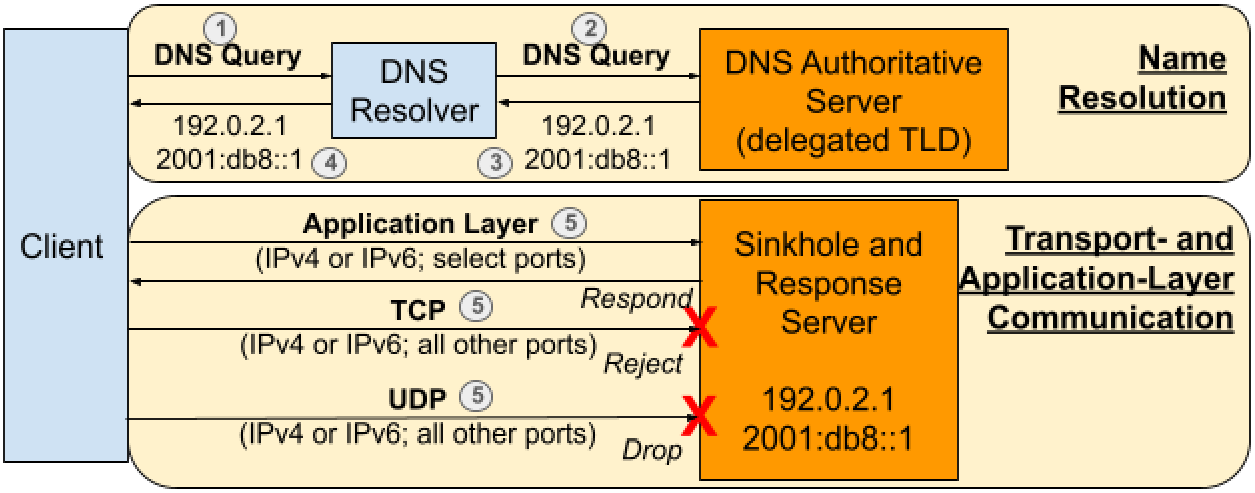

Client
1
DNS Query
DNS Resolver
2
DNS Query
DNS Authoritative Server (delegated TLD)
Name Resolution
192.0.2.1
2001:db8::1
4
3
192.0.2.1
2001:db8::1
Application Layer
5
(IPv4 or IPv6; select ports)
Respond
TCP
5
(IPv4 or IPv6; all other ports)
Reject
UDP
5
(IPv4 or IPv6; all other ports)
Drop
Sinkhole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

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.[71]

As the DG is neither in a position to assess such non-technical risks nor to operationalize mitigation strategies, the DG looks to ICANN to implement the relevant recommendations and necessary procedures required to limit potential negative impacts to the DNS and the ICANN org.

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

4 Findings

After reviewing years of earlier work, including the Interisle Report[72], the JAS Report[73], 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.[74] 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

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

[73] 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

[74] 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

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."[75]

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."[76] Even with that finding, 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."

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

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.[77]

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.

4.2.1 Name collisions continue to persist within the DNS

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

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

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

the examples of .CORP, .HOME, and .MAIL, name collisions continue to occur even ten years after their original identification as name 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.[78] 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.

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.

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

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.[79] This data can only assist with providing a leading indicator of potential impact. Determining the harm

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

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 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.[80]

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.[81]

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 name 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.

"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."[82]

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

[80] 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."

[81] 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/.

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

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.

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.[83] 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.

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

[83] 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

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."[84] 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."[85] 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 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.

[84] 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

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

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.")[86]

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

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."[87] 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

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

[87] 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

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.[88] 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 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.[89]

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 mechanism to notify the global recursive resolvers or others who may find that information necessary for their operations. No process exists to signal to global public resolvers when they need to obtain new copies of root zone data out of their typical schedule.

[88] See IANA Update to the GNSO, October 2023, https://static.sched.com/hosted_files/icann78/ef/iana-icann78-gnso-update-202310.pdf
[89] See "Root Zone Maintainer Agreement (RZMA)" - ICANN, https://www.icann.org/iana_imp_docs/129-root-zone-maintainer-service-agreement-v-28sep16.

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[90] 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.[91] 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.[92]

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

As noted in the JAS Report[93], several of the NCAP DG findings, and SAC 113: SSAC Advisory on Private-Use TLDs[94], 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

[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 "Statistics," Google IPv6, https://www.google.com/intl/en/ipv6/statistics.html

[92] See ICANN's IPv6 Initiative, https://www.icann.org/resources/pages/ipv6-initiative-2017-02-28-en

[93] 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

[94] 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.

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.[95] 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.

[95] 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").[96]

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? [97]

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.

[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

[97] 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 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.

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

The Root Cause Analysis Report 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.[98] 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]."[99] 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.

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

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

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

[99] 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

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 4.4.2: 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[100] and the New gTLD Collision Management Proposal[101] regarding mitigation measures that can be taken by using methods similar to "trial delegations.".

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.

[100] 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

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

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.

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

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.

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.

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).

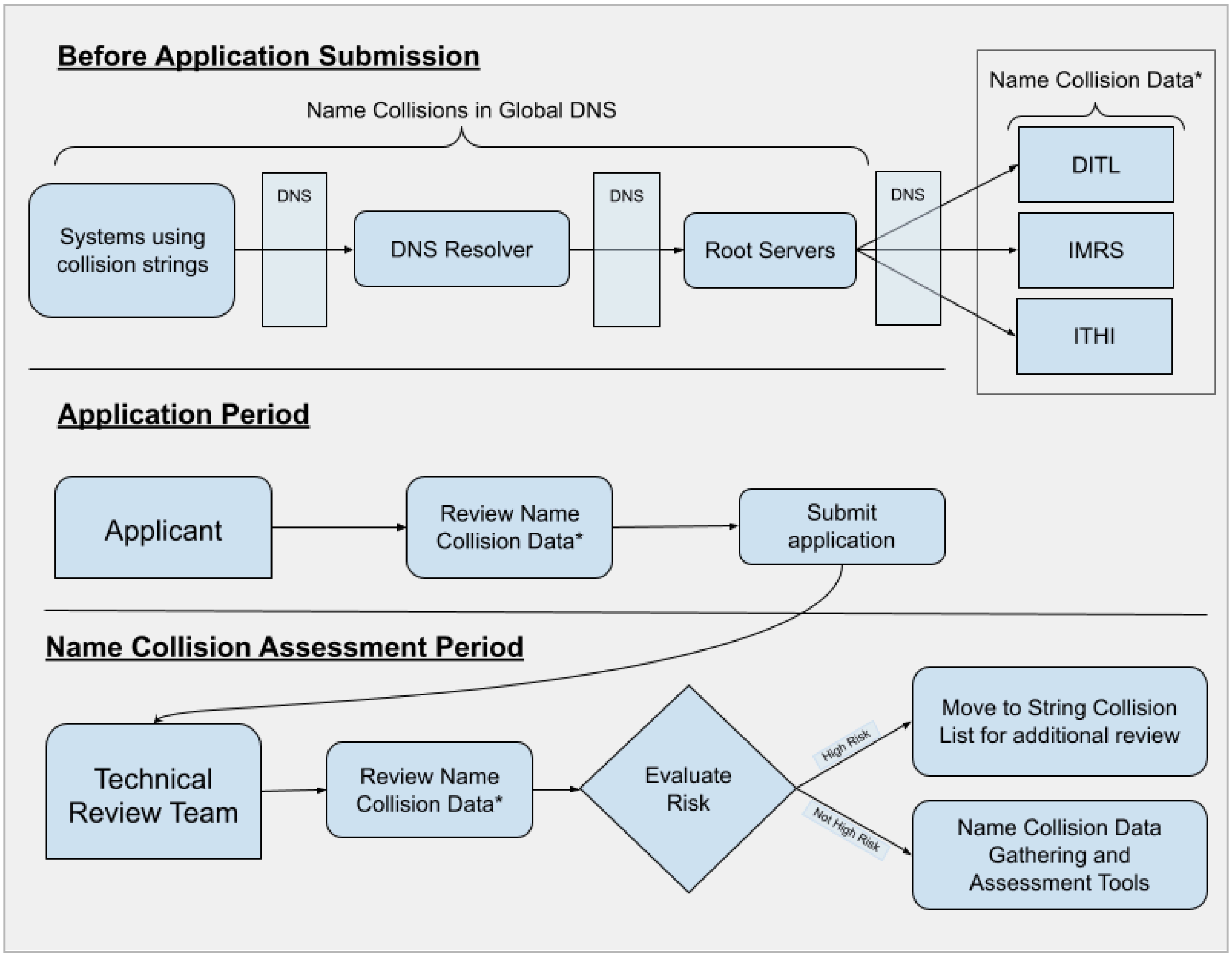

Before Application Submission
Name Collisions in Global DNS
Name Collision Data*
Systems using collision strings
DNS
DNS Resolver
DNS
Root Servers
DNS
DITL
IMRS
ITHI
Application Period
Applicant
Review Name Collision Data*
Submit application
Name Collision Assessment Period
Technical Review Team
Review 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.

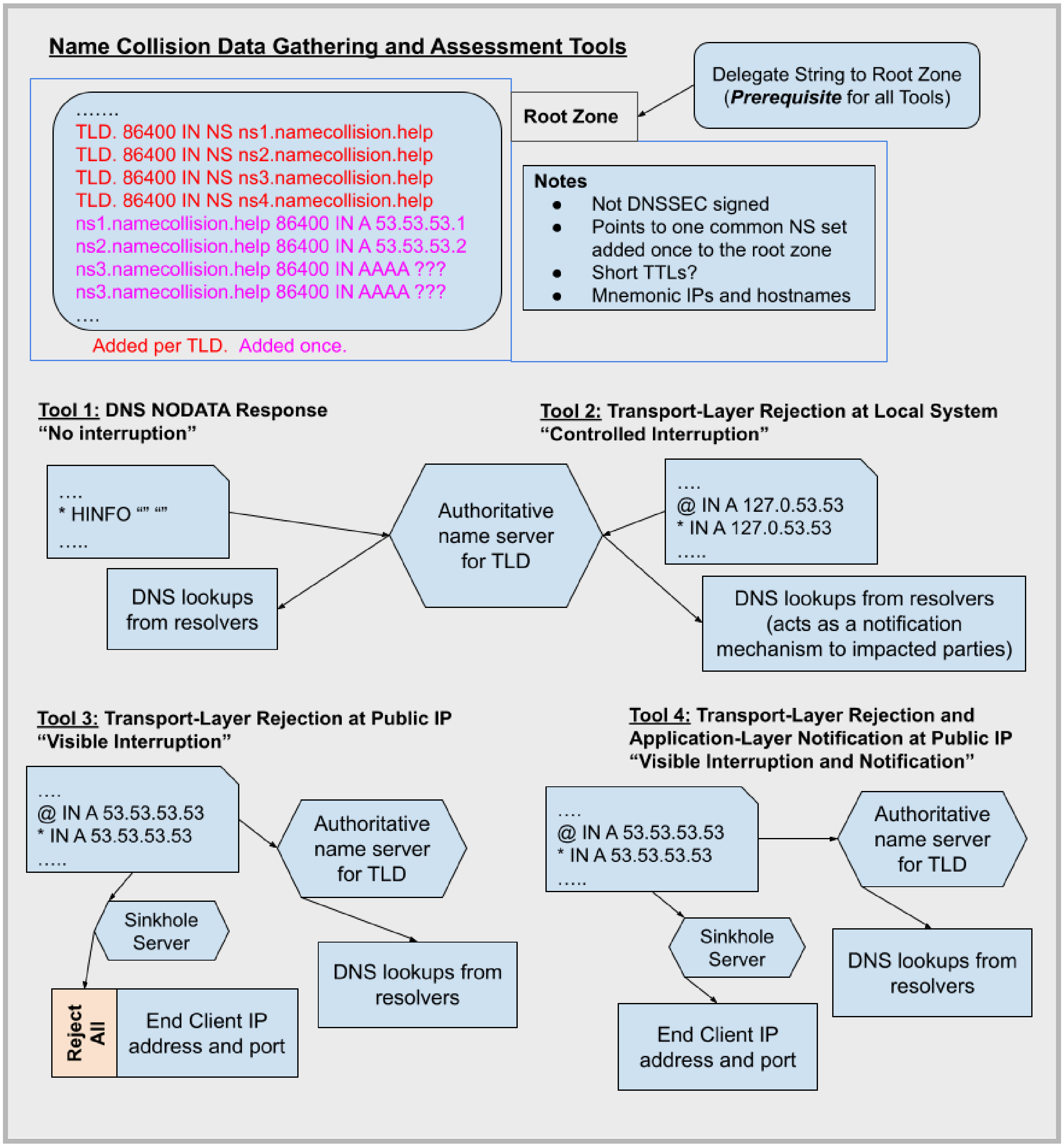

Name Collision Data Gathering and Assessment Tools
…….
TLD. 86400 IN NS ns1.namecollision.help
TLD. 86400 IN NS ns2.namecollision.help
TLD. 86400 IN NS ns3.namecollision.help
TLD. 86400 IN NS ns4.namecollision.help
ns1.namecollision.help 86400 IN A 53.53.53.1
ns2.namecollision.help 86400 IN A 53.53.53.2
ns3.namecollision.help 86400 IN AAAA ???
ns3.namecollision.help 86400 IN AAAA ???
….
Added per TLD. Added once.
Root Zone
Delegate String to Root Zone (Prerequisite for all Tools)
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 "" ""
…..
Authoritative name server for TLD
DNS lookups from resolvers
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
DNS lookups from resolvers
Reject All
End Client IP address and port
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
DNS lookups from resolvers
End Client IP address and port

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, though any data collected should be aggregated and anonymized 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 name 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 name 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.[102] 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 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.

[102] "Report a Name Collision," ICANN, accessed 17 January 2024, https://www.icann.org/en/forms/report-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[103].

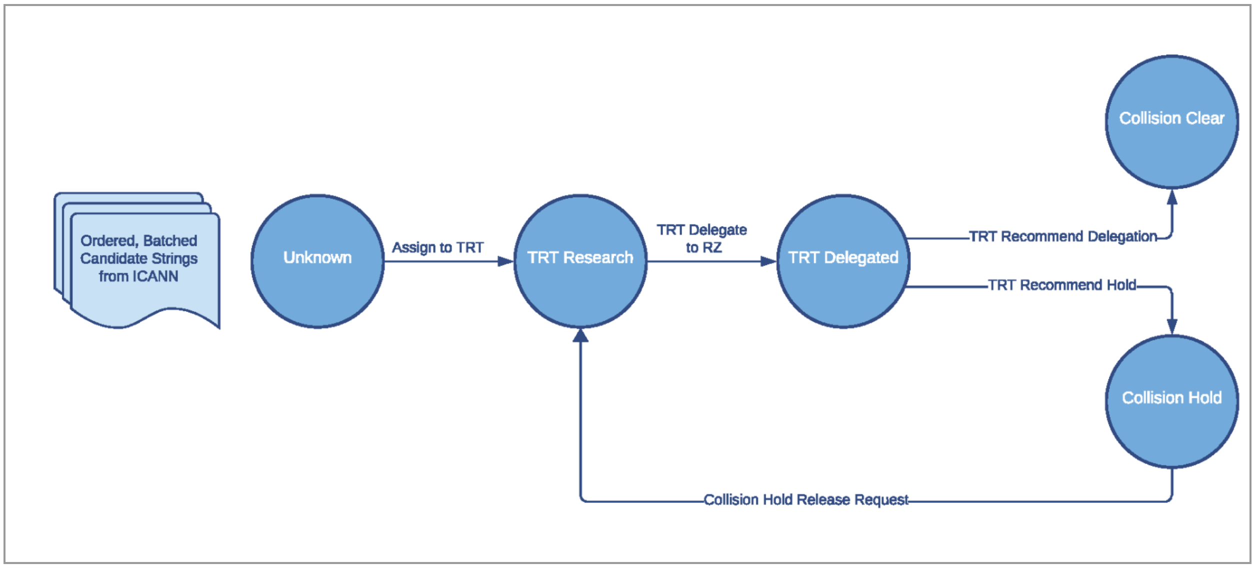


Figure 8: Representation of Technical Review Team workflow for assessing strings

[103] 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

5.9.1 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).[104] 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.

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 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

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

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 request the emergency removal of one of the strings they have delegated as part of the Name Collision Risk Assessment Framework. They should, however, 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”[105] 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”[106]:

> “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.

[105] 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

[106] 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[107] 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.

[107] 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[108].

[108] 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[109] 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.

[109] 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, as well as a timeline that laid out what everyone might expect from a name collision risk assessment process.

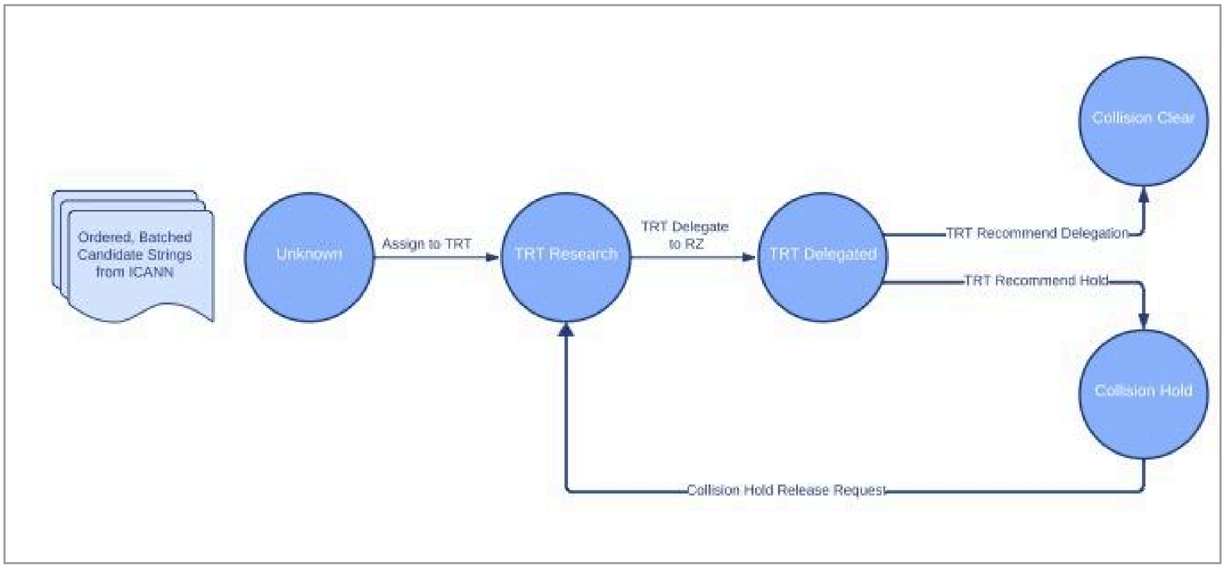


Figure 9: Representation of Technical Review Team workflow for assessing strings

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 three points in time during the application process: during static assessment, during Passive Collision Assessment, and during Active Collision Assessment. At each point, the TRT is expected to document their results as part of making a recommendation to move onto the next assessment activity (i.e., moving from static to passive to active assessment).

During static assessment, 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 Passive Collision Assessment, 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 the root cause of the name collision. 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. During Active Collision Assessment, if undertaken, the TRT will 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).

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.

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. While no such process exists today for the emergency removal of a delegation, the DG determined this is a natural and necessary part of the assessment workflow.

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 details of that workflow can be found in Recommendation 8. 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.

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; ICANN has no such process at this time.

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.

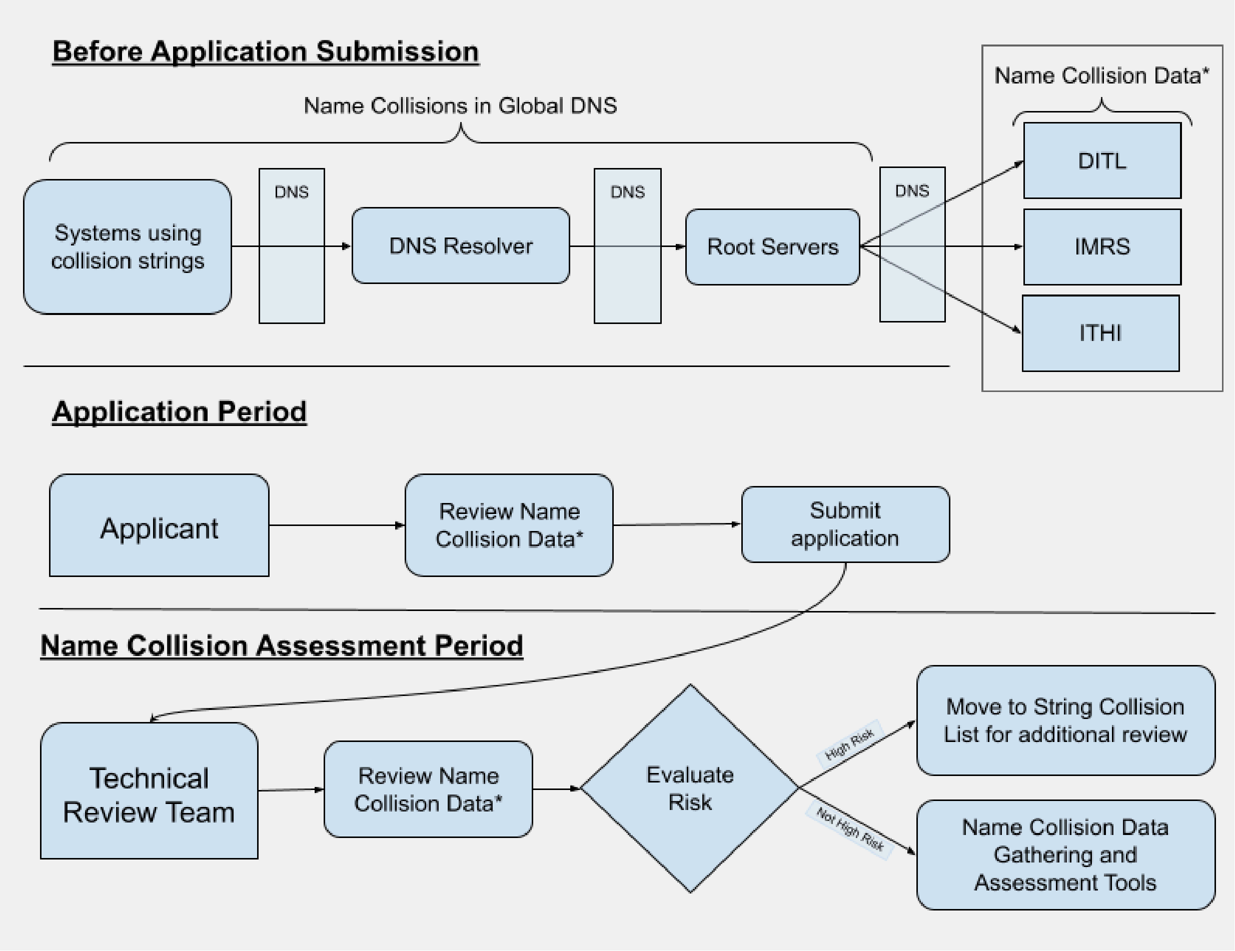
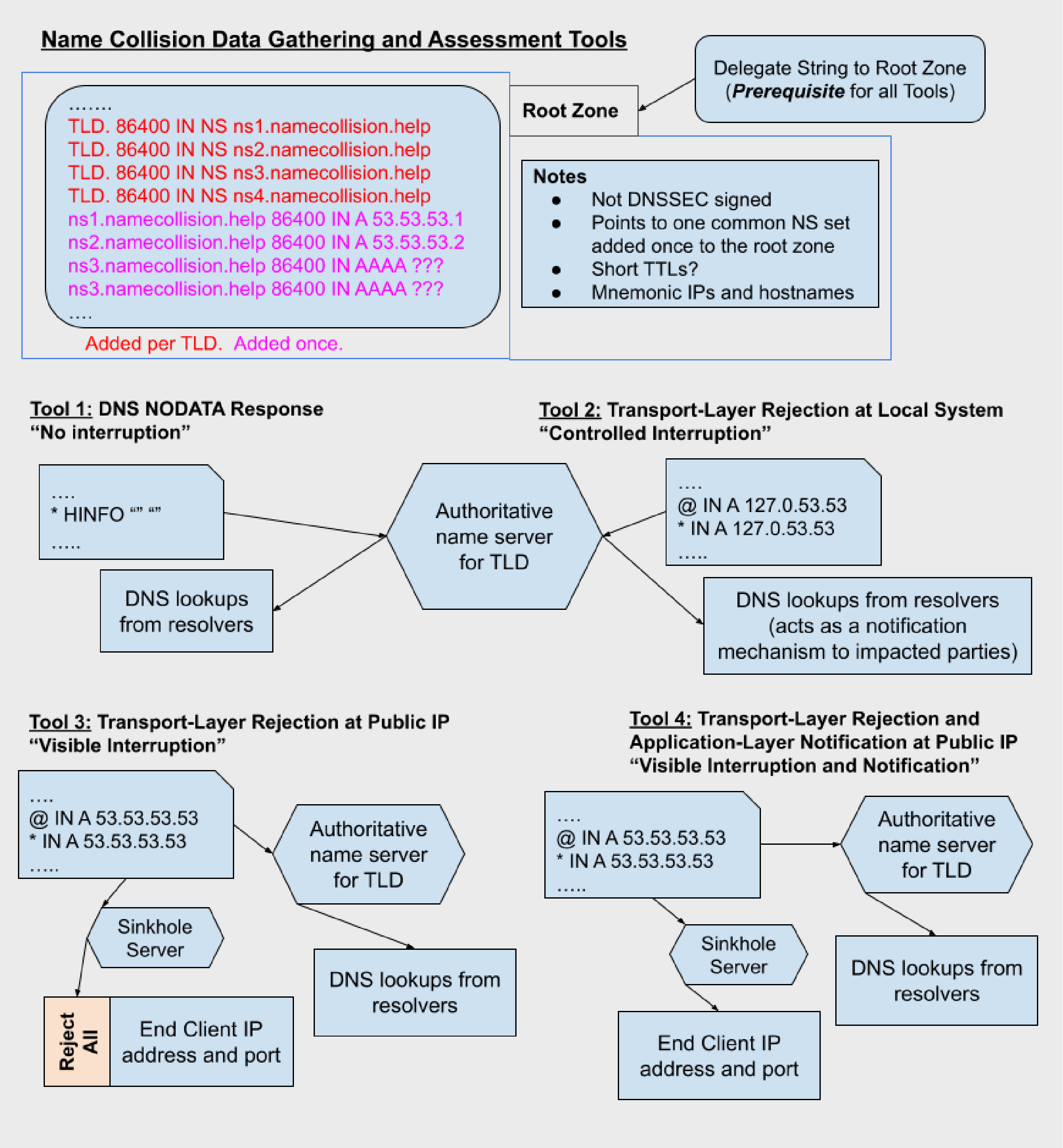

Before Application Submission
Name Collisions in Global DNS
Systems using collision strings
DNS
DNS Resolver
DNS
Root Servers
DNS
Name Collision Data*
DITL
IMRS
ITHI
Application Period
Applicant
Review Name Collision Data*
Submit application
Name Collision Assessment Period
Technical Review Team
Review 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