DRAFT SSAC/NCAP Study 2 Report

A Report from the Security and Stability Advisory Committee

DD Month YYYY

Preface

In this document the Security and Stability Advisory Committee (SSAC) {TBD}.

The SSAC focuses on matters relating to the security and integrity of the Internet's naming and address allocation systems. This includes operational matters (e.g., pertaining to the correct and reliable operation of the root zone publication system), technical administration matters (e.g., pertaining to address allocation and Internet number assignment), and registration matters (e.g., pertaining to registry and registrar services). SSAC engages in ongoing threat assessment and risk analysis of the Internet naming and address allocation services to assess where the principal threats to stability and security lie, and advises the ICANN community accordingly. The SSAC has no authority to regulate, enforce, or adjudicate. Those functions belong to other parties, and the advice offered here should be evaluated on its merits. SSAC members participate as individuals, not as representatives of their employers or other organizations. SSAC consensus on a document occurs when the listed authors agree on the content and recommendations with no final objections from the remainder of the SSAC, with the exception of any dissenting opinions or alternative views that are included at the end of the document.

Table of Contents

Executive Summary

TBA

1 Introduction

The Name Collision Analysis Project (NCAP) Study 2 final report brings together the research and analysis of several studies that touch key facets around the issue of name collision. The report is structured in such a way as to walk the reader through the methodology and findings of the three research studies. The conclusions from those studies in turn supports the answers to the questions the ICANN Board has asked the NCAP to respond to and the recommendations on how to proceed when handling the potential for name collision in the DNS.

This first section describes the background of the NCAP and the mandate set forth by the ICANN Board in 2017. It goes on to describe the methodology of the study group as a whole, including the timeline of research, community outreach, and study group consensus. The study group found several points where they needed to establish a working group definition for several terms; the results of those discussions is a terminology section that makes clear how those terms are used in this report.

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, may find the material relevant to their efforts.

1.1 Background and Related Work

The work behind the NCAP began a decade ago with SAC057: SSAC Advisory on Internal Name Certificates, wherein the SSAC referred to the issue of "name collision" and provided the ICANN Board with steps for mitigating the issue. On 18 May 2013, the ICANN Board adopted resolution 2013.05.18.09 – 2013.05.18.11, regarding SAC057, commissioning a study on the use of TLDs that are not currently delegated at the root level of the public DNS in enterprises. From there, the work continued to evolve as the understanding regarding the depth and breadth of the issue grew.[1]

In 2014, ICANN published the "Mitigating the Risk of DNS Namespace Collisions Final Report," a report by JAS Global Advisors ("JAS") (the final report was published in 2015).[2] This report was used to develop the “Name Collision Occurrence Management Framework” to guide ICANN and the new gTLD registry operators on how to handle name collisions.[3]

Moving ahead to 2017, the ICANN Board requested via resolutions (2017.11.02.29 - 2017.11.02.31) that the ICANN Security and Stability Advisory Committee (SSAC) conduct studies to present data, analysis and points of view, and provide advice to the Board on the topics

[1] “History of the Name Collision Analysis Project,” https://community.icann.org/display/NCAP/History+of+the+Name++Collision+Analysis+Project.

[2] https://www.icann.org/en/announcements/details/mitigating-the-risk-of-dns-namespace-collisions-final-report-by-jas-global-advisors-30-11-2015-en

[3] https://www.icann.org/en/system/files/files/name-collision-framework-30jul14-en.pdf

around DNS name collision. In response, SSAC formed the Name Collision Analysis Project.[4] This project is organized into three studies. The first study, which provided a primer on the topic of name collision and a list of datasets that either existed at the time of the study or would need to be generated to support further analysis, was finalized on 19 June 2020 after submission to the Board and a public consultation period.

Whereas Study 1 established baseline documentation, the scope of Study 2 was revised based on new information regarding the DNS (e.g., the use of new transport protocols for the DNS, the existence of new data sets). The revised scope was described in "SSAC2021-02: Revised Study Two Proposal for the Name Collision Analysis Project" and accepted as per resolutions 2021.03.25.11 – 2021.03.25.14. The revised scope focused on four key areas:

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

The third study, to understand the ramifications of name collisions with an ultimate goal of developing concrete guidance on how to avoid negative consequences, will be informed by the work from studies one and two.

1.2 Methodology

With the acceptance of the revised Study 2 proposal, the discussion group kicked off the study reports as described in Section 3. Study Reports and settled into a regular meeting cadence. While the discussion group considered the questions assigned by the ICANN Board, the researchers collected and analyzed available data relevant to understanding both how to observe and how to measure the impact of name collision. Each effort required coordination to make sure the Board questions were supported by the findings of the study reports, and that the study reports stayed in scope with the overall mandate for the group.

ICANN org provided administrative support for the NCAP, including project management and technical writing resources. [ICANN org also funded the research analyst(s)]

The discussion group chairs called for consensus on the responses to the Board questions, the study reports, and any special terminology after discussion on each item was concluded during the regular conference calls.

Throughout the NCAP study efforts, the study group has considered collisions at level below the TLD level and alternate roots as out of scope.

[4] "Invitation: Name Collision Analysis Project (NCAP) Discussion Group," https://www.icann.org/en/announcements/details/invitation-name-collision-analysis-project-ncap-discussion-group-17-4-2019-en.

[All study reports went out for public consultation prior to their being used in this report to finalize the findings and recommendations to the ICANN Board.]

Item	Final Consensus Call Date	Result
Case Study of Collision Strings		[Full|Partial|None]
Perspective Study of DNS Queries for Non-Existent Top-Level Domains (was Data Sensitivity Analysis		[Full|Partial|None]
Root Cause Analysis		[Full|Partial|None]
Final Report		[Full|Partial|None]

1.3 Terminology

- Allocation - The process by which the Board makes a decision about allowing a name to be delegated.
- Collision Strings - (from the Proposed Definition of Name Collisions and Scope of Inquiry for the Name Collisions Analysis Project) a string that manifests name collisions
    - Collision String Registry
    - reservation - a registry of names not to be allocated nor delegated (on the collision string list). Note that multiple organizations (i.e., IETF, ICANN) can add names to the reserved list.
- Controlled Interruption - (From FAQ) "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."
- 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 initially undertaken every year since 2006. This data has historically been the primary measurement asset for name collision studies.
- Delegation - This process may occur after allocation; it describes the technical process of adding an allocated name to the DNS root zone.
- 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.
- Name Collision - (used in Study 1 and RFP) Name collision "refers to the situation where a name that is defined and used in one namespace may also appear in another. Users and applications intending to use a name in one namespace may attempt to use it in a different

one, and unexpected behavior may result where the intended use of the name is not the same in both namespaces. The circumstances that lead to a name collision could be accidental or malicious."
    - Domain Name Collision - A name collision in the single resolution protocol of the DNS.
    - Name Collision Assessment - Controlled Interruption is a mechanism for Name Collision Assessment
    - Namespace Collision - A potential source of name collision involving multiple namespaces, such as the DNS root zone and a blockchain service.
- Query Volume - The number of DNS requests received for a string.
- Root Server Identity (RSI) - thirteen identities, each of which is named with the letters 'a' to 'm', collectively administered by twelve root server operators. They are authoritative for the 'root-server.net' domain.
- 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." [5]
- 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

2 Advice to the ICANN Board

["provide advice to the Board regarding the risks posed to users and end systems if .CORP, .HOME, .MAIL strings were to be delegated in the root, as well as possible courses of action that might mitigate the identified risks." – Does our advice go beyond this original mandate re: just those gTLDs? Are we saying that our advice is generally applicable?]
[What exactly will ICANN need to do? Tool development to support active collision assessment? Process changes? Policy changes?]

3 Study Reports

As described in its revised scope, the NCAP Study Group 2 conducted three studies:

- Case Study of Collision Strings
- A Perspective Study of DNS Queries for Non-Existent Top-Level Domains
- Root Cause Analysis

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

[5] https://www.icann.org/en/system/files/files/sac-064-en.pdf

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 of the overall picture of the DNS queries that would help identify name collisions.

The third and last report, the Root Cause Analysis, considered known name collisions and evaluated what mitigation or remediation happened, particularly what they did and why. The following sections describe the results of those studies in greater detail; the full report for each is available in Annexes A through C at the conclusion of this report.

3.1 Case Study of Collision Strings

The NCAP discussion group met over the course of approximately two years to evaluate and consider questions posed by the ICANN Board on the delegation of currently reserved TLDs such as .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. Among these include the observation that queries for these undelegated names are both increasing in volume and diversity. These facts suggest that challenges relating to impact and mitigation 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.

[Hypothesis/Goal: 1) defined all the critical diagnostic measurements, gave them a name and created the set. For us, name collisions are defined by the existence of those measurements. 2) to shift from focusing on harm to focusing on the impact of name collisions. We came up with guidance on how impact can be assessed (two characteristics of the CDM: volume as an absolute number, diversity of that number). More volume in CDMs and more diversity within those CDMs drives the impact.]

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

[Hypotheses/Goals: two things: any root server operator is representative of root server operations. That makes the fact that ICANN is already publishing visible name collisions out of L-root and making that publicly available. Second, resolver operators see a different view of the DNS than root server operators. Right now, Matt only has one global resolver, would like to have others, but we've already learned the most important thing (that they are different from RSOs). That they are different drives the controlled interruption and the requirement for it. Because you need controlled interruption to pull data out of resolvers so root servers can see it. You have to do

the delegation to do the controlled interruption and see any conflicts. Also, will never see the resolver data again, it's not public. It's not so much about the spread as it is about the pockets of individuals. It comes with search suffixes wrt how organizations are configured. So many places do this in-house that you won't see it at the global level. There can be further analysis on this, and it's not analysis we're exactly doing right now. The root server data will never provide you a guaranteed view of all name collisions.]

3.3 Root Cause Analysis

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

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

4 Board Questions

Part of Board resolution 2017.11.02.30 were a set of topics that helped define much of the scope for the discussion group. While the resolution referred to these as questions, they were not in question format. The discussion group found it a valuable exercise to reconsider each topic as a question; the responses to each are below.

4.1 Defining Name Collision

> *(1) a proper definition for name collision and the underlying reasons why strings that manifest name collisions are so heavily used;*

The term "name collison" has been defined in several formal documents. For example, NCAP Study 1 states:

> For the purposes of Study 1, the definition of "name collision" specified by the NCAP Discussion Group (DG) for the RFP was used. Name collision "refers to the situation

where a name that is defined and used in one namespace may also appear in another. Users and applications intending to use a name in one namespace may attempt to use it in a different one, and unexpected behavior may result where the intended use of the name is not the same in both namespaces. The circumstances that lead to a name collision could be accidental or malicious.

Study 1 concerns name collisions in the context of top-level domains (TLDs), where the conflicting namespaces are:

- the global Internet Domain Name System (DNS) namespace reflected in the root zone overseen by the Internet Assigned Numbers Authority (IANA) Function; and
- any other namespace, regardless of whether that other namespace is intended for use with the DNS or any other protocol."

The original RFP for Study 1 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."

[Add clarifying text about namespace collision as a term we need to be aware of. When we get to remediation/mitigation, a source of some collisions is outside ICANN's DNS namespace. So you might get r/m plans that attempt to address that. We won't spend time studying this in any depth. And if this is the only place where we use the term, we probably don't need it in terminology.]

4.2 Negative Answers

Board topic	Question as understood by the NCAP DG
(2) the role that negative answers currently returned from queries to the root for these strings play in the experience of the end user, including in the operation of existing end systems;	What is the role that negative answers currently returned from queries to the root for these strings play in the experience of the end user, including in the operation of existing end systems?

As noted in the SSAC Report, "Redirection in the Com and Net Domains," uninstantiated names that result in negative answers might occur for a variety of reasons: "A name might not exist

because it had been misspelled, had lapsed or had never been registered. A name might also be registered or reserved but not included in the lookup database used for domain name queries."[6] Regardless of the reason, the errors received when returning a negative answer are in and of themselves both useful to systems and end users and a vector for harm. For example, systems such as spam filtering services may rely on the error to help determine if a message is spam by checking whether the domain name of the sender exists. Alternatively, any interruption or intervention in the path that results in a negative answer has the potential to intrude upon end-user privacy by allowing the intervening system to collect data on the user's behavior and the path attempted.[7] From a system perspective, interruption or intervention in the flow by a third party could result in increased network charges for some classes of users, a reduction in performance, or the creation of work required to compensate for the consequent failure.[8]

Search lists, one prominent form of signal interruption in the DNS that is configured per host, may also introduce unintended behavior through negative answers. As per the SSAC Advisory on Search List Processing, search list behavior is not standardized (see Table 1 in SAC 064) but "is a contributor to the invalid queries seen at the root servers."[9] The SSAC study reports that .home and .corp as the two most common labels for negative answers (NXDOMAIN results) based on analysis of data collected by the Domain Name System Operations, Analysis, and Research Center (DNS-OARC) and reported by the Day in The Life of the Internet (DITL) project. The label .mail is also on the list.

When a host processes its search list in response to a negative answer for the first search term, it may attempt different suffixes or try an unqualified instead of a qualified domain name. As covered at length in SAC064, the rules are complex and inconsistent from host to host. When search lists were first developed, the root zone was a static entity; that assumption is built into the very design of search lists. The design that used second, third, and lower labels in a domain name (such as CORP, HOME, and MAIL) would never appear in the root and so would be easily resolved with search list processing. This assumption is no longer valid and the result is a highly variable, inconsistent behavior that the average end user will not understand.

[Search list processing is dependent on NXDOMAIN. The DNS is just one example of an app/service which is dependent on NXDOMAIN queries working a certain way. The DNS is an integral part of the Internet infrastructure. It's the worst case scenario of what SAC 064 is getting at. It's not that the DNS depends on NXDOMAIN queries, it's that the expected behavior for the configuration is dependent on certain things not resolving.

[6] pg 3
[7] pg 22
[8] pg 23
[9] SAC064 pg 11

See as an example of the Google Chromium behavior - it used NXDOMAIN probes up to the root to see if DNS interception was occurring. That was an application behavior logic that expected certain behavior of NXDOMAIN. If the suffix is suddenly delegated, the signaling would change and so the associated behavior would change.

This applies to other scenarios where DNS queries are being made and the response becomes inconsistent and may even be weaponized [DNS-SD, WPAD, ISATAP, etc.].
A negative response is a technical protocol element. The SSAC report suggest there are two broad categories of why negative answers occur: accidental, which is always going to be a thing; and inadvertent, such that whatever the user is interacting with, whatever process the user is interactive with, the negative answer is required in order for things to work the way they would expect. They play a role when a user is doing something that needs that negative answer in order to provide the user something. We have two examples: search lists and application-specific behavior. The former is trying to provide a consistent experience to users with what's valid and what's not, and the latter is trying to protect the user. Both of these things bear out that there is an intentional use of NXDOMAIN that the user might not directly know about, but the application/service the user is using is aware of that functionality and is taking advantage of it. We can't know in advance all the innovations that might happen in the future; these are examples of what happened in the past. See the case study for corp home and mail for how bad it got. If the UX changes from a way the user expects it to, to something else, you open up a threat vector (see section 4.3 Harm)]

-

4.3 Harm

Board topic	Question as understood by the NCAP DG
(3) the harm to existing users that may occur if Collision Strings were to be delegated, including harm due to end systems no longer receiving a negative response and additional potential harm if the delegated registry accidentally or purposely exploited subsequent queries from these end systems, and any other types of harm;	What are the types of harm and their likelihood to existing users if Collision Strings were to be delegated? This should include considerations around harm due to end systems no longer receiving a negative response and additional potential harm if the delegated registry accidentally or purposely exploited subsequent queries from these end systems, as well as any other types of harm.
(4) possible courses of action that might mitigate harm;	What possible courses of action can ICANN org take that might mitigate harm?

(5) factors that affect potential success of the courses of actions to mitigate harm;	

Regarding the question of "harm," the discussion group focused on the potential for harm. The connotation of "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. Real-world, demonstrable harm has been difficult to identify due to the limitations in both available data and the lack of clarity around the definition of harm. Where the discussion group found evidence of actual, clear harm, we have called that out specifically.

The responses to the board questions are not intended to address (a) the probability of an end user being harmed in any of these manners, (b) the frequency with which these harms would occur, (c) the degree of harm (if any) that could be incurred by any particular end user, (d) whether the existing end system's intended to leverage negative answers in designing its systems, nor (e) whether such harm could be avoided or mitigated in ways other than refusing to delegate strings. Instead, we have focused on offering information that will help the board with their decisions regarding domains on, or even yet to be added, to the collision string list.

Harms may come through several possible threat vectors. While these vectors are not specific to the DNS, these vectors may open as a result of name collisions. For example, a name collision may result in a user unexpectedly arriving at a service where they may see false information or be required to submit credentials or disclose other private information that will be harvested for nefarious purposes (i.e., a phishing attack). While harms are difficult to directly identify, the threat vectors that allow harm to occur are more easily identified and mitigated. Mitigating the threat vectors logically mitigates the potential for harm. More information about threat vectors that impact the DNS are available in the DNS Security Facilitation Initiative Technical Study Group's (DSFI-TSG) final report to the ICANN CEO.[10]

For the sake of this discussion, we have categorized the controllable threat vectors that open the door to the potential for harm as involving either Interception and Manipulation or Signaling Interruption.

Interception and Manipulation includes an attacker intercepting DNS queries and either answering the queries directly or changing, manipulating, or providing false answers. These would be queries that were previously negatively answered by the root servers and can subsequently be received and answered by various parties, either purposefully or unknowingly, after the delegation of a TLD string. In such a scenario, an attacker's exploitation of name collisions will allow them to intercept and manipulate DNS queries. Through these name

[10] DNS Security Facilitation Initiative Technical Study Group (DSFI-TSG), "DNS Security Facilitation Initiative Technical Study Group - Final Report, October 2021, https://community.icann.org/display/DSFI/DSFI+TSG+Final+Report.

collision events, attackers may capitalize on a variety of attack vectors as noted in the DSFI-TSG report.[11]

Signaling Interruption, as mentioned in Board Question 2, discusses the role played by negative answers currently returned from queries to the root. This could include breaking applications that utilize the DNS as a signaling tool rather than as a directory (e.g. Chrome startup, Mozilla DoH, etc.). Search list processing, as described in Section 4.2 is an example of mainstream signal interruption. Another scenario is one in which conditional logic of the returned DNS answer is baked into the application and can be handled in many different ways. Unfortunately, the scale of variations makes it impossible with current technologies to measure, assess, or remediate this potential for harm.

Mitigating name collisions is particularly difficult as detecting and reporting name collisions is itself challenging. At the local level, organizations are unlikely to be able to see the problem (e.g. transient corporate devices used on corporate networks) or even be able to reliably trace the causes. On a broader level, registrars and registries are unlikely to detect name collisions until well after the fact. As described further in Section 6. General Recommendations for this report, there will need to be a workflow that supports collecting information and having the ICANN Board evaluate each case individually to determine the appropriate course of action based on their analysis of the potential risk.

4.4 Risks of Delegation

Board topic	Question as understood by the NCAP DG
(6) potential residual risks of delegating Collision Strings even after taking actions to mitigate harm;	What are the potential residual risks of delegating Collision Strings even after taking the actions described in Board Question 4 to mitigate harm?

4.5 Undelegated Strings and Collision Strings

[11] Some of these attack vectors and corresponding risks stem from DNS-SD or zero-configuration protocols that utilize the DNS as a bootstrapping mechanism. When coupling those protocols with either intentional rooting of a namespace in an undelegated TLD or through unintended consequences of suffix search lists, these types of queries are often the most exploitable attack vector in a name collision scenario.

Board topic	Question as understood by the NCAP DG
(7) suggested criteria for determining whether an undelegated string should be considered a string that manifest name collisions, (i.e.) placed in the category of a Collision String;	What are the suggested criteria for determining whether an undelegated string should be considered a string that manifests name collisions, (i.e.) placed in the category of a Collision String?
(8) suggested criteria for determining whether a Collision String should not be delegated, and suggested criteria for determining how remove an undelegated string from the list of Collision Strings; and	What are the suggested criteria for determining when a collision has been sufficiently mitigated that a Collision String can be removed from the list.
(9) measures to protect against intentional or unintentional creation of situations, such as queries for undelegated strings, which might cause such strings to be placed in a Collision String category, and research into risk of possible negative effects, if any, of creation of such a collision string list.	What measures would be appropriate and effective to protect against intentional or unintentional creation of situations that might cause strings to be placed in a Collision String category? What are the potential negative effects of a collision string list?

By "Collision String", the NCAP Discussion Group assumes the Board is asking for a method of identifying a Top Level string that, due to the high risk of name collisions associated with the string and the potential for harm (as described in response to Board Question #3), that particular string should not be delegated and should be reserved. The DG notes that there is, in fact, a spectrum or "range" of risks of harm associated with delegation of each new string as was observed in connection with ICANN's Alternate Path to Delegation process followed in the 2012 round, which permitted more rapid delegation of certain strings provided certain names were "blocked" at the second level. In this regard, the subsequent Name Collision Framework which adopted a system of 90-day "controlled interruption" is a system which lends itself to identification of name collisions, but does not, in and of itself, mitigate those collisions.

5 Analysis and Findings

The study reports described above provided a wealth of information that informed both the responses to the board questions and the general recommendations below.

With the Case Study of Collision Strings, we learned what information is needed to identify and define the critical diagnostic measurements required to identify name collisions. In particular, we were able to understand how to evaluate the level of impact as determined by the diversity and volume of queries. With the data at hand, we can see that while name collisions remain an issue,

there is data that will allow for the development of appropriate mitigation and/or remediation strategies. We can also show that the telemetry data used by an RSI will continue to diminish in its fidelity as DNS protocol changes (e.g., QNAME minimization) further get deployed in the DNS.

The Perspective Study of DNS Queries for Non-Existent Top-Level Domains demonstrated that the available data sets from any one RSI is representative of the overall picture of the DNS queries at RSIs that would help identify name collisions. Global resolvers, however have their own views that are not publicly visible. In practice, this means that ICANN org can rely on the data it has available via the L RSI to give a leading indicator to the community of some name collisions risks, but additional data is required to surface the data currently only available in the global public resolvers. In order to make that data visible and be certain of the existence of name collision activity, the TLD must be delegated to the root zone.

[The third and last report, the Root Cause Analysis, considered known name collisions and evaluated what mitigation or remediation happened, particularly what they did and why. The following sections describe the results of those studies in greater detail; the full report for each is available in Annexes A through C at the conclusion of this report.]

With the conclusions coming from the Case Study and Perspective Study, we demonstrate that Controlled Interruption provides one way to capture whether or not a name collision exists by using the data from any of the RSIs and to send a signal to the client that there was a collision. It does this, however, by co-opting certain behaviors of the DNS protocol: it replaces the NXDOMAIN (non-existent domain) response with an A (address record) response that will surprise a client and result in unknown failure modes. While this is intended to be a graceful failure, this mechanism introduces a potentially negative impact on the client.

An alternative approach that would allow for an initial assessment to verify the existence of a name collision is Passive Collision Assessment. The NCAP Study Group identifies this as a way to capture whether or not a domain collision exists by delegating the TLD to the root zone, without sending the “magic” IP address back to the client.

Controlled Interruption vs Passive Collision Assessment

In the 2012 New gTLD Round, Controlled Interruption was defined as “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.” The implementation deployed was to delegate the approved TLD string into the root zone and add a wildcard address record into the TLD zone that returned an unroutable “magic” IP address, i.e., 127.0.53.53, that would presumably motivate any client receiving it to investigate.[12] This was

[12] From the JAS report, page 15: “Because the primary objective is to communicate with system administrators through their logs, this unique and strange IP should stand out in log files, be noticed, and result in the administrator searching the Internet for assistance (we note that as of today, using Google to search for “127.0.53.53,” the top 5 results are relevant).” https://www.icann.org/en/system/files/files/name-collision-mitigation-study-06jun14-en.pdf

intended to make name collisions generally visible but also introduced a risk of unexpected impact on the client.

We propose establishing a new mechanism, **Passive Collision Assessment,** that ensures the existence of name collisions is visible at root servers and is limited by the period of time during which their existence is observed. Passive Collision Assessment differs from Controlled Interruption in the following ways:

	Passive Collision Assessment	**Controlled Interruption**
Primary goal	to make name collisions visible to the RSI	to make name collisions visible to the RSI and notify the client of a name collision
Mechanism	delegate the proposed TLD with no additional labels to the root zone	delegate the proposed TLD to the root zone, including a wildcard A-record
Notification	there is no notification to the client	there is a notification in the form of a special IP address of 127.0.53.53 back to the client
Risks	no known risks to client as DNS protocol is not disrupted	potential risk of impact to the client as a result of disrupted DNS protocol behavior

As of the writing of this document, this minimum requirement of establishing a mechanism to make it possible to observe the existence of a name collision is easily met by adding the proposed TLD string to the root zone and ensuring no other labels are added to the TLD zone. This will ensure that any requests for domain names will continue to receive responses with content equivalent to what was received prior to adding the proposed TLD string to the root zone.

Observing the existence of a name collision is not sufficient, however, to allow for a full risk analysis as to whether that domain can be delegated without harm to others. It is a necessary first step, but once a name collision has been identified, additional data such as query volume and query type diversity (see Critical Diagnostic Measurements) are required to investigate and execute a technical analysis and develop a plan for mitigation or remediation.

Critical Diagnostic Measurements

As highlighted in the Case Study report, 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.[13] The different measurements must be taken as a whole to understand how their interactions inform any technical analysis. For example,

> "query volume--one of the four major classes of measurements--is an important factor, but a single source that could be easily mitigated with a simple configuration 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."[14]

The four major classes of measurement, in no particular order, include:[15]

- Query Origin Diversity - the number of unique query source IP addresses (resolvers)
- Label Diversity - diversity of labels under a name collision string
- Query Volume - the number of queries each RSI receives
- Query Type Diversity - the type of query (i.e., resource record type) being requested.

Understanding Enhanced Controlled Interruption

Neither Controlled Interruption nor Passive Collision Measurement allows for evaluating the full impact of the name collision. They do not allow for capturing the diversity nor volume of queries when they come in through protocols outside the DNS, two of the Critical Diagnostic Measures for understanding name collisions. Understanding the level of impact is a critical component when evaluating the risk of delegating a given string. To get at that level of information, there is a need to collect additional information. One mechanism for that is Active Collision Assessment as it allows for reviewing data that comes through other protocols.

[Active Collision Assessment is one mechanism to collect additional data regarding a name collision. Unlike Controlled Interruption, ECI returns a routable IP address of a system that would then provide a protocol-appropriate response. [How does it do this?] While this meets the need to understand the diversity of name collisions across additional protocols (e.g., DoH), there are also significant concerns regarding the potential for such a system to end up collecting personal or private data from a client.] - We're making clear the requirements of what data needs to be collected, in order to develop a mitigation/remediation plan.

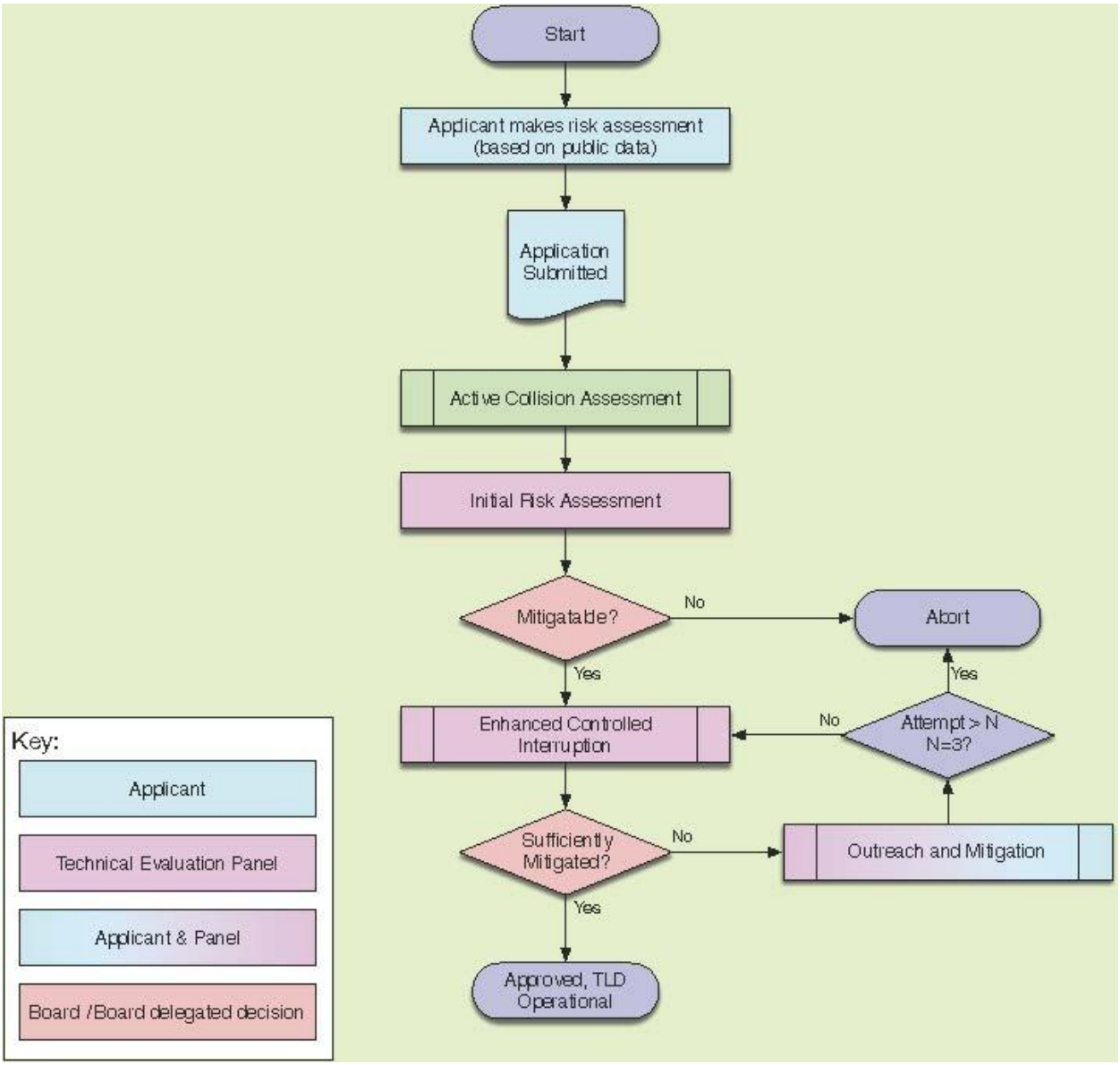
---

[13] Case Study, pg 26
[14] Case Study, pg 27
[15] Open-source Intelligence (OSINT) strings are also mentioned in the Case Study, but those strings 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.

6 General Recommendations

[Paper will already have told ICANN to prepare to change their tools/process/etc. Here's where we walk them through the details of the workflow with a focus on what applicants will need to do.]
[Technical Review Team: there's a function that has to happen. Coming out of the case study we make the observation that the critical diagnostic measurements exist and you're looking for certain criteria. You're looking for a technical evaluation about what's coming out of controlled interruption. We label that role as a "technical review team" - this is a function that has to happen at this point in time. We describe all the things this function has to do, but we don't speak to how that function is implemented. It could be a third-party contract, it could be a group within OCTO, or some other model that the board needs to decide on. – maybe not have this here]

6.1 Application Workflow



6.2 Board Decision Process

[Board will consider the input and recommendation of the Technical Review Team and make a decision.]

7 Conclusion

[NCAP Study 3?]

We need controlled interruption

- Resolvers see collisions that roots don’t
- We won’t see resolver data again

We need enhanced controlled interruption

- To investigate the origin of the name collision
- Protocol source - listen on all ports
- Protocol data - what’s on the inside to assess the activity
- QNAME minimization
- DoH/DoT/DoQ??
- Aggressive NSEC Caching, NXDomain Cut

8 Acknowledgments, Statements of Interest, and Dissents, Alternative Views and Withdrawals

In the interest of transparency, these sections provide the reader with information about aspects of the SSAC process. The Acknowledgments section lists the SSAC members, outside experts, and ICANN staff who contributed directly to this particular document. The Statements of Interest section points to the biographies of all SSAC members, which disclose any interests that might represent a conflict—real, apparent, or potential—with a member's participation in the preparation of this Report. The Dissents and Alternative Views section provides a place for individual members to describe any disagreement with, or alternative view of, the content of this document or the process for preparing it. The Withdrawals section identifies individual members who have recused themselves from discussion of the topic with which this report is concerned. Except for members listed in either the Dissents and Alternative Views or Withdrawals sections, this document has the consensus approval of all of the members of SSAC.

8.1 Acknowledgments

The committee wishes to thank the following SSAC members and experts for their time, contributions, and review in producing this report.

SSAC Members
{TBD}

NCAP Discussion Group Members
{TBD}

ICANN staff
{TBD}

8.2 Statements of Interest

SSAC member biographical information and Statements of Interest are available at:
https://www.icann.org/resources/pages/ssac-biographies-2021-10-21-en

NCAP Discussion Group member Disclosure of Interest are available at:
https://community.icann.org/display/NCAP/NCAP+Discussion+Group

8.3 Dissents and Alternative Views

Appendix A: Table of Board Question vs Study Two

Directly copied from "SSAC2021-02: Revised Study Two Proposal for the Name Collision Analysis Project"[16]

Board Questions	Study Two Tasks
(1) a proper definition for name collision and the underlying reasons why strings that manifest name collisions are so heavily used;	Completed during Study One but subject to revision according to analysis in Study Two
(2) the role that negative answers currently returned from queries to the root for these strings play in the experience of the end user, including in the operation of existing end systems;	Conduct impact analysis
(3) the harm to existing users that may occur if Collision Strings were to be delegated, including harm due to end systems no longer receiving a negative response and additional potential harm if the delegated registry accidentally or purposely exploited subsequent queries from these end systems, and any other types of harm;	Conduct root cause analysis Conduct impact analysis
(4) possible courses of action that might mitigate harm;	Conduct root cause analysis Conduct impact analysis ● Study Three Tasks to follow
(5) factors that affect potential success of the courses of actions to mitigate harm;	● Study Three Tasks to follow
(6) potential residual risks of delegating Collision Strings even after taking actions to mitigate harm;	Conduct impact analysis ● Study Three Tasks to follow
(7) suggested criteria for determining whether an undelegated string should be considered a string that manifest name collisions, (i.e.) placed in the category of a Collision String;	Produce a report on the results of Study Two

[16] https://www.icann.org/groups/ssac/documents-correspondence

(8) suggested criteria for determining whether a Collision String should not be delegated, and suggested criteria for determining how remove an undelegated string from the list of Collision Strings; and	Produce a report on the results of Study Two
(9) measures to protect against intentional or unintentional creation of situations, such as queries for undelegated strings, which might cause such strings to be placed in a Collision String category, and research into risk of possible negative effects, if any, of creation of such a collision string list.	Produce a report on the results of Study Two ● Study Three Tasks to follow
(10) to present data, analysis and points of view, and provide advice to the Board regarding the risks posed to users and end systems if .CORP, .HOME, .MAIL strings were to be delegated in the root, as well as possible courses of action that might mitigate the identified risks.	Produce a report on the results of Study Two

Appendix B: Case Study of Collision Strings

Appendix C: A Perspective Study of DNS Queries for Non-Existent Top-Level Domains

Appendix D: Root Cause Analysis