DRAFT SSAC NCAP DG Responses to Board resolution 2017.11.02.30

DD Month YYYY

Table of Contents

Responses to Board resolution 2017.11.02.30

On 25 March 2021, the ICANN Board passed Resolution 2021.03.25.11 – 2021.03.25.14 directing the Name Collision Analysis Project Discussion Group (NCAP DG) to proceed with Study Two as redesigned by SSAC 2021-02: Revised Study Two Proposal for the Name Collision Analysis Project (5 February 2021).[1] The revised proposal modified the original expectations of NCAP Study Two such that it removed two of the original goals, "Building a data repository" and "Build a test system which can be used for impact analysis and to test possible mitigation strategies." The revised proposal also shifted most of the work slated for paid contractors to the group itself. Overall, the results of these modifications reduced the scope, level of effort, total costs, and resources to execute Study Two.

As part of Resolution 2021.03.25.13, the Board reinforced "the continued relevance of the nine questions related to name collisions presented in Board resolutions 2017.11.02.29 - 2017.11.02.31, especially questions (7) and (8) concerning criteria for identifying collision strings and determining if collision strings are safe to be delegated."[2]

The topics covered in Board resolution 2017.11.02.30 initially defined the structure and activities of the NCAP DG.[3] As the group considered each topic, we found that members had different interpretations of what the Board was expecting in response to the resolutions. Rather than debate the Board's expectations, the discussion group found it a valuable exercise to reconsider each topic as a question and focus on providing a considered, thoughtful response.

This document categorizes the nine questions into six key areas:

- Defining Name Collision (question 1)
- Negative Answers (question 2)
- Harm (question 3)
- Mitigating Harm (questions 4 and 5)
- Risks of Delegation (question 6)
- Undelegated Strings and Collision Strings (questions 7, 8, and 9)

[1] https://www.icann.org/en/board-activities-and-meetings/materials/approved-resolutions-regular-meeting-of-the-icann-board-25-03-2021-en#2.b and https://www.icann.org/en/system/files/files/ssac2021-02-05feb21-en.pdf

[2] https://www.icann.org/en/board-activities-and-meetings/materials/approved-resolutions-regular-meeting-of-the-icann-board-25-03-2021-en#2.b

[3] https://www.icann.org/en/board-activities-and-meetings/materials/approved-resolutions-regular-meeting-of-the-icann-board-02-11-2017-en#2.a

Defining Name Collision

Board topic	Question as understood by the NCAP DG
(1) a proper definition for name collision and the underlying reasons why strings that manifest name collisions are so heavily used;	What is the full definition of the term 'name collision'? What are the underlying reasons why strings that manifest name collisions are so heavily used?

The term "name collision" has been defined in slightly different ways across several formal documents.[4] In order to reach a consistent and clear definition per Board question 1, the NCAP DG offers a recommendation for definition of name collision. This definition maps to one sent out for public comment prior to starting NCAP Study 1.[5] The NCAP DG endorses the following definition: ~~name collision~~

> "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 a formal denotation of name collisions is provided in the background section of the Study Two Final Report.[6]

Clearly describing what constitutes a name collision is a necessary step to identifying the underlying reasons behind why they occur. Previous research conducted by JAS Global Advisors established a taxonomy that led to an understanding that "(1) very few root causes seem to explain the vast majority of colliding behavior, and (2) nearly all root causes appear in all TLDs in differing proportions. Only .corp, .home, and .mail are clear outliers."[7] The taxonomy consists of six classifications and is thoroughly described in the JAS report. The NCAP Case

[4] See https://www.icann.org/resources/pages/name-collision-2013-12-06-en, Interisle report (https://www.icann.org/en/announcements/details/addressing-the-consequences-of-name-collisions-5-8-2013-en), NCAP Study One

[5] ICANN org, "Proposed Definition of Name Collisions and Scope of Inquiry for the Name Collisions Analysis Project," 19 August 2019, 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

[6] TBA

[7] https://www.icann.org/en/system/files/files/name-collision-mitigation-final-28oct15-en.pdf pg 34

Study of Corp, Home, and Mail also reaffirms these findings and highlights similar root cause reasons in other high query volume name collision TLDs.

The NCAP Study Two revised proposal included the following:

> “Using the similar data sources and methodologies by JAS Global Advisors and Interisle Consulting Group, perform updated case studies of the CORP, MAIL, HOME, and other strings. The study should highlight changes over time of the properties of DNS queries, and traffic alterations as a result of DNS evolution.”[8]

As a result of that research, several possibilities were identified as potential causes for name collisions. Actual causes, however, for TLD-level name collisions are contained in the NCAP Root Cause Analysis, identified as Study 2 Task 1.[9] The Root Cause Analysis examines the documented name collision occurrences reported to ICANN as well as incidents found in Web search results.[10] Per the Root Cause Analysis, the origins of name collisions were diverse, both in terms of the application involved and their root causes. Multiple applications were involved, some that users interfaced with directly and others that were more process-driven. In terms of the domains used, they were found in both private and non-private namespace, using both fully-qualified and unqualified domain names (including unqualified names with single and those with multiple labels).

Furthermore, the Root Cause Analysis found that the private use of DNS suffixes is widespread. It is clear from the data that the private use of DNS suffixes is not isolated. Private use of DNS suffixes is exhibited within over half of newly delegated TLDs, even though a few TLDs are responsible for more usage than others.

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 role do the negative answers currently returned from queries to the root for these strings play in the end user's experience, including any experience in the operation of existing end systems?

[8] https://www.icann.org/en/system/files/files/ssac2021-02-05feb21-en.pdf pg 6
[9] https://www.icann.org/en/system/files/files/ssac2021-02-05feb21-en.pdf pg 5
[10] TBA

As noted in the SSAC Report, “Redirection in the Com and Net Domains,” uninstantiated names that result in negative answers might occur for various 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.”[11] Ultimately, enumerating all possible ramifications of negative answers on end users and applications is not possible; every application may react differently to negative answers. Those reactions ultimately depend on whatever signal is used internally to indicate a name does not exist.

Regardless of the reason, the errors received when returning a negative answer are both useful to systems and end users. 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 sender’s domain name exists. Alternatively, any change from a negative answer to a routable and serviceable IP address 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.[12] 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.[13]

Another example of how negative answers have a role in system behavior and the end-user experience is the search list.[14] Hosts commonly use search lists for facilitating the resolution of names within common DNS suffixes used by that host. As covered at length in SAC064, the rules are complex and inconsistent from host to host.[15] Given a name to be resolved, the host may iterate through the different suffixes in the list or try an unqualified instead of a qualified domain name, depending on the outcome of the previous iterative resolution attempt. Often the resolution outcome expected by the user relies on one or more previous resolution attempts resulting in negative answers.

Search lists have particular relevance to TLD-related name collisions when they involve the private use of suffixes under TLDs that have not previously been delegated. When search lists were first developed, the root zone was a static entity. Search list design that used second, third, and lower labels in a domain name (such as CORP, HOME, and MAIL) that would never appear in the root and would be easily resolved with search list processing because of negative responses. This assumption is no longer valid. When a search list includes TLDs that exist or may exist in the public DNS, the end-user experience might be highly variable, ranging from no disruption to complete interruption, without a clear understanding of the cause.

[11] https://www.icann.org/en/committees/security/ssac-report-09jul04.pdf pg 3

[12] pg 22

[13] pg 23

[14] ICANN Security and Stability Advisory Committee, “SAC064: SSAC Advisory on DNS “Search List” Processing,” report, 13 February 2014, https://www.icann.org/en/system/files/files/sac-064-en.pdf.

[15] ibid; see Section 6

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.

To address the Board's question, the discussion group focused on two aspects of harm: potential harm and reported 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.

Potential Harm
We have identified three general categories of potential harm related to name collisions: DNS Query Surveillance, Communication Disruption, and Communication Interception. We describe each category in increasing order of potential harm.

DNS Query Surveillance. Some portion of leaked DNS queries for domain names under undelegated TLDs has always reached the root servers. However, once those TLDs are delegated, some fraction of them reach not only the root servers but also the servers authoritative for the TLDs—and possibly other servers as well. This contact allows additional parties to monitor incoming DNS queries that were most likely not even intended to be exposed to the public Internet. The harm that might be felt by users experiencing this behavior is a function of 1) the nature of the domain names being leaked, 2) the extent to which those queries are being logged and monitored, 3) the relationship between the authoritative servers and the users or organizations from which the queries originate. In the most innocuous sense, the nature of the queries might be inconsequential, the authoritative servers oblivious, and the organizations without any substantial relationship. However, in a more severe case, queries might be ultra-sensitive, revealing secret or embarrassing information, the query logs actively monitored by operators, and the servers in cahoots with adversaries of the user or organization. This data leakage could result in a loss of reputation, public embarrassment, or even costly lawsuits.

Communication Disruption. After the delegation of a TLD, queries for names under that TLD might yield an IP address (or other positive response appropriate for the query type) where they previously did not. Such a positive response might reach an end system where a *negative* response is expected. The very fact that the response contains an answer might disrupt certain applications, services, or entire networks. This is typically due to one of the following: 1) the positive response short-circuited a resolution process that would have produced the *expected* answer, or 2) ultimately, the resolution process was not expected to produce *any* answer at all. In either case, subsequent application behavior typically involves communicating with the address returned, contrary to expectations. This disruption might affect not only the application itself but also other dependent applications. Harm, in this case, might be quantified by estimating the time and other resources dedicated to identifying the root cause of the disruption and remediating the problem by adjusting network configurations, individual system configurations, or user behaviors. This solution might be trivial for an individual or a small organization but relatively complicated and expensive for a large organization. However, the diversity of systems that depend on the Internet makes both the systems themselves and the potential for harm, in the case of name collision, difficult to identify and assess.

Communication Interception. When an application receives an unexpected positive response from the public DNS, the application potentially attempts communications with the entity associated with the IP address. In the case of communication disruption, the communication is rejected or goes unanswered, either because the IP address is unreachable or there is no service responding on the port in question. However, if the IP address is reachable and responsive, the outcome is communication interception. In the case where the service exists, but the content returned is identified by the user or system as unexpected, the behavior and content would provide a basis for investigation. A more harmful scenario is when the content returned is intended to impersonate legitimate content, with the objective of obtaining sensitive information, such as credentials or proprietary information. While the first scenario is likely accidental, the second is related to explicit exploit attempts. Harm in these cases ranges from that associated with disruption to loss of sensitive information.

While these threats are described separately, a user can experience harm in more than one of these categories simultaneously. For example, active surveillance might lead to intentional interception.

Reported Harm

While identifying potential harm is useful in understanding the variety of ways systems might be affected by name collisions, a review of reported harm validates that potential and highlights the most significant instances.

Two sources of data for assessing and quantifying harm experienced with name collisions are the set of name collisions reports submitted to ICANN and the responses to name collisions surveys, further described and analyzed in the Root Cause Analysis. All reports and survey responses can be categorized as communications disruption directly related to controlled interruption. There are no reported instances of DNS query surveillance or communication interception, and no reports present evidence of any collisions from circumstances other than controlled interruption.

The impact of the incident that prompted each name collision report submitted was categorized as either severe, significant, small-scale, or unknown, based on the number of users or systems that were affected, the number of applications that were affected, or other subjective detail provided in the report.[16] Half of the reports indicated experiencing severe (21%) or significant (29%) impact. Reports involving severe impact included the following comments: “more [than] 30,000 employees in over 7 countries”, “Network down, no internet access”, and “The scale of the impact is fairly critical”. Reports involving significant impact included the following: “150 users”, “Unable to send mail”, and “No network shares access”. Nonetheless, we note that *no* report indicated any “clear and present danger to human life”—which text was provided as a condition for submission on the Web submission form.

Similarly, one of the survey respondents made the following comment: “This was very expensive and disruptive. In addition, employees cannot reach websites in the network domain.”[17]

Summary

In summary, we have described harm as the potential negative impact that might be felt by individuals and organizations experiencing name collisions, and we have listed specific instances of this impact that have come from the root cause analysis. We note that all reported instances of harm thus far can be categorized as communication disruption and can be directly traced to controlled interruption. However, the primary purpose of this disruption is to alert users and prompt configuration and behavior changes to avoid future name collisions that might lead to more severe harm, e.g., communication interception with the intent to exploit.

Mitigating Harm

Board topic	Question as understood by the NCAP DG
(4) possible courses of action that might mitigate harm;	What possible courses of action can ICANN org take that might mitigate harm?

[16] TBA

[17] TBA

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

The need to mitigate harm implies the presence of harm. However, possible courses of action involve not only mitigating harm but also reducing the likelihood that a negative impact is felt at all. We describe both categories of action, including the parties that might be expected to take action in each case.

Reactive Measures to Mitigate Harm

Action by ICANN. The most extreme action that ICANN org can take to mitigate harm associated with the delegation of a TLD is the **removal of its delegation**. The JAS report considers this option "feasible [but] undesirable as it creates considerable opportunity for operational complexities and unintended consequences." (p11). The same report opines that "de-delegation of a TLD in the root would effectively be a permanent death for that TLD" (p11). Other actions that ICANN org might take include the following:

- **Provide a means whereby parties negatively impacted by name collisions can report their experience.** The name collisions report form is an example of this. The reports submitted to that form provide one of the few qualitative data sources with which we can assess the impact of name collisions. However, the current text on the form introduces a bias in the data because individuals are deterred from submitting a report unless "your system is suffering demonstrably severe harm … or you have a reasonable belief that the name collision presents a clear and present danger to human life".[18] Less restrictive text would allow greater insights into the harms of name collisions and possibly suggest additional courses of action.
- **Offer technical assistance to parties negatively impacted by name collisions.** While interactive and/or individual technical support might not be feasible (support which the JAS report deems out-of-scope for ICANN org), making general resources available for technical self-help is a completely reasonable course of action.[19] This is especially true considering the abundance of knowledge of root causes identified and analyzed in the Root Cause Analysis Report, the NCAP Study 1 Report, the JAS report, and other studies.
- **Refer affected parties to the registry associated with the TLD at the heart of the name collisions for further action.** In at least one case, action was taken by one registry because ICANN org acted on a report submitted through the form.[20]

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

[19] JAS report, pg 10

[20] Study 1 Report, p37

Action by Registry. When it is known that name collisions are causing harm, the registry also has courses of action. One of the most extreme actions that might be taken is **removing the delegation of a second-level domain from the zone**. In the case of controlled interruption, the equivalent action is introducing a temporary exception to the wildcard record in place for that domain (see "Implementation Guidance 29.6" in the SubPro Final Report). There is already precedent for this type of action; Study 1 reports that in one case,

> "a large organization had reported disruption of its services on the first day after new TLD delegation. The registry operator for the new TLD voluntarily chose to temporarily stop controlled interruption for that TLD. After the affected organization updated its systems to correct the problem, the registry operator was able to resume controlled interruption for the TLD".[21]

Another course of action by a registry is to **offer technical assistance to parties negatively impacted by name collisions**. While interactive and/or individual technical support might not be feasible (support which the JAS report deems out-of-scope for registries), making general resources available for technical self-help is a completely reasonable course of action.[22] Just as with similar resources that might be provided by ICANN org, there is a wealth of knowledge related to name collision root causes from previous studies. The value of having resources at the registry level, independent of resources provided by ICANN org, is two-fold: (1) there might be TLD-specific technical nuances (e.g., public configuration examples that use the TLD in private naming context) that are most appropriately made available by the registry; and (2) the registry and registrar are and registrar are more closely associated with the registrant than ICANN org is and the registry or registrar could provide additional contextualized assistance to the impacted parties or registrant.and the registry or registrar could provide additional contextualized assistance to the impacted parties or registrant..

Proactive Measures to Reduce the Potential for Harm

Controlled interruption is one of the measures ICANN org and contracted parties have implemented with the intent to reduce the potential for harm. The goal of controlled interruption is to alert systems that might experience harm from name collisions *in the future*, in the hopes that administrators will discover the problem and implement changes in configuration and/or behavior that reduce or eliminate the likelihood of *future harm*. However, the very disruptions that make this alerting effective often cause harm themselves. The justification for this is that the near-term harm is inflicted with good intentions by a knowledgeable entity, the mechanism is contained within a finite period of time (90 days from delegation), and it does not involve the exchange of any application-layer data. In contrast, longer-term harm might be caused either accidentally by an unknown party or maliciously by a knowledgeable entity, the timing is

[21] (Study 1 Report, p37)

[22] JAS report, pg 10

completely unknown, and application-layer data might be exchanged. Thus, controlled interruption is analogous to getting a vaccination that possibly causes immediate, short-term illness to prevent a possible worse illness in the future.

The Root Cause Analysis shares data related to the questions of the near-term harm associated with controlled interruption (the only harm that we know about thus far) and the possible longer-term harm. In the Root Cause Analysis, survey data shows that 70% of respondents that used private namespace experienced problems related to controlled interruption. Of the reports submitted to ICANN org via their name collisions form, half suggested that the impact felt by controlled interruption was either significant or severe. However, the Root Cause Analysis document also shows that new mappings (i.e., to non-controlled interruption IP addresses) were introduced for names within 20% of domains and 28% of TLDs that were observed to have experienced name collisions, all within 18 months of delegation. While this alone does not imply a long-term name collision, it does indicate that there is potential.

Another way of proactively reducing the likelihood of harm before it occurs is by **analyzing DNS query logs for behaviors indicating that name collisions might result should a TLD be delegated**. Characteristics such as high volume and/or high diversity of queries and query names under a given TLD provide some indicator that a non-existent TLD string is in active use. These characteristics alone cannot definitively confirm nor quantify the potential for name collisions, just as their absence cannot definitively confirm a lack of collision potential. Nonetheless, the Root Cause Analysis has shown a correlation between these metrics and actual reported name collisions and harm. TLD strings with relatively high metrics can be the basis for reaching out to potentially affected parties, understanding configurations, and preemptively encouraging them to modify their configuration, particularly if a TLD is being applied for or is soon to be delegated. This exercise should be repeated periodically because network devices, configurations, and behaviors will change over time, resulting in changes to the problem set of TLD strings.

Verisign has already used this proactive approach in at least two separate efforts.[23] First, using query logs for two of the root servers, A-root and J-root, they identified 46 TLD strings with the potential for name collisions and possible related harm. They communicated with administrators of affected parties to inform, identify the root cause, and facilitate configuration change. In all cases, the underlying parties confirmed the problem. One of the primary root causes involved the use of suffix search lists to resolve unqualified domain names. Changes were instituted in the case of roughly half of the organizations contacted, the effects of which were observed within a few months by way of dramatic decreases in query metrics. In other cases, change is expected to take a long time because of the time required to develop and (especially) deploy firmware in devices that have a large user base. While none of the 46 TLD strings were applied for in the

[23] https://blog.verisign.com/domain-names/verisign-outreach-program-remediates-billions-of-name-collision-queries/

2012 round of new gTLDs, this proactive effort reduced the likelihood of name collisions and possible harm, should they be applied for and delegated in the future. Additionally, the configuration changes reduced the inadvertent leaking of private and sometimes sensitive DNS data into the public global DNS.

Verisign also conducted investigations regarding the TLD string, CBA.[24] While ultimately no root cause was confirmed in this case, and the string was delegated, it is another example of due diligence.

In some cases, proactive investigation of name collisions might yield a set of TLD strings whose query characteristics are significantly high, enough so that outreach to identify the root cause and encourage configuration change might prove to be infeasible—at least in the short term. In such cases, it might be more prudent to **maintain a collision string registry of potentially problematic strings** (this is discussed in the section "Undelegated Strings and Collision Strings" later in this paper). The presence of a TLD string on such a list would effectively prohibit it from being delegated until such time as the potential for harm could be thoroughly investigated, root causes identified, and problematic configurations addressed or cleared. Both the outreach mitigation action and the identification of problematic name collision strings can be done in advance and independently of any TLD-application round. Both proactive mitigations help prevent harm to existing systems and networks, future applicants, and ICANN org.

While proactive measures can be successful in reducing the likelihood of harm associated with name collision, the effectiveness of proactive efforts is dependent on the ability to collect data, the data's completeness and robustness, the ability to analyze and distill such data, the ability to correlate the name collision traffic or data with impacted parties, networks, and services, the outreach efforts, and the cooperation of affected parties.

Summary

Both reactive and proactive measures can be taken to mitigate the potential for harm associated with name collisions. Both can be effective techniques, but both also have limitations, such that the potential for harm cannot be completely eliminated.

Risks of Delegation

Board topic	Question as understood by the NCAP DG

[24] https://www.verisign.com/assets/report-cba-analysis.pdf

(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?

It is important to note that there will always be some risk associated with the delegation of new TLD strings, particularly those that have been identified as collision strings (see question 7). While the techniques proposed for both reducing the likelihood of potential harm and mitigating harm (question 4) reflect due diligence, the following facts remain:

- We are limited to the data we have available to make assessments with regard to name collisions;
- The data itself has limitations with respect to its visibility and what can be inferred from the analysis thereof;
- Quantitative assessments are only a heuristic for measuring the impact associated with name collisions and might not accurately reflect; and
- Behaviors and configurations might change from those currently employed, introducing name collisions for which there was previously only *potential*.

Thus, whether because of incomplete data, imperfect assessments of data, or future, unforeseen changes, the risk of harm associated with delegation of a collision string, or even a string that does not currently manifest name collisions, is non-zero.

Undelegated Strings and Collision Strings

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, or, in other words, is 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 of Collision Strings?

(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?

Criteria for Identifying Collision Strings

Applying the definition of name collision from Board question 1, we see the practical manifestation of name collisions associated with undelegated TLD strings as the case where a user or software produces a DNS query that is *not intended* to reach the RZM root servers but *does* reach an RZM root server—and is answered by that server. This query might be considered a "leak" from the alternate namespace. When a "name error" (i.e., "NXDOMAIN" response code) response is returned by the root server, the user or software continues resolution in its own namespace, uninterrupted. In this case, the name collision is *observable* at the RZM root server but *undetected* by and thus *immaterial* to the user or software. However, name collisions become *material* when non-NXDOMAIN responses are received from the RZM root servers, i.e., post-delegation.

An undelegated TLD string that manifests name collisions (i.e., a "Collision String") will result in DNS queries that can be observed at the RZM root servers. Thus, the presence of queries at RZM root servers for names under a given undelegated TLD string is often used to *infer* name collisions for that domain. However, without knowing the root cause or origin of the queries, a *definitive* statement attributing them to name collision risk or harm cannot be made. Nevertheless, undelegated TLD strings for which queries are observed at the RZM root servers are *treated* as if they manifest name collisions—unless it can be proven otherwise. The attention given to these undelegated strings is roughly proportional to the volume and diversity of the queries observed; without further qualitative data, it is assumed that higher query volume and higher query diversity suggest that a larger population of users and software is affected.

An undelegated string can be locally suppressed or answered by an entity's internal DNS infrastructure and thus would never be observed at the RZM root servers. Without perfect visibility into the entire DNS ecosystem, there is no deterministic way to tell if a string manifests name collisions; therefore, all undelegated strings should be considered as having collision potential but are not yet name collision strings.

Criteria for Determining Whether a Collision String Should Not Be Delegated

When an undelegated TLD string exhibits query metrics that are *significantly higher* than those of other undelegated strings, there is reason to believe that the string is being used by many users or systems, possibly due to a widespread software implementation or configuration. For example, the JAS report identified such metrics with the strings `corp`, `home`, and `mail`[25], and the Case Study report additionally showed that their use is growing, as indicated by increasingly high metrics.[26] The JAS report recommended that they "be referred to the Internet Engineering Task Force (IETF) for potential RFC 1918-like protection/treatment" in that they are not delegated—equivalent to a list of "Collision String[s] that should not be delegated". Just as with collision strings in general, candidate strings for the "do-not-delegate list" should not be definitively classified as such without additional qualitative data, but they should be treated as such until their impact can be proven otherwise through qualitative means.

Independent of the query metrics associated with an undelegated TLD string, if there is reason to believe that the delegation of that string would result in "severe harm" such as would present "a clear and present danger to human life", then that string should not be delegated.[27]

Criteria for Determining the Removal of a Collision String from the Do-Not-Delegate List
To qualify for removal from the list of strings that should not be delegated, the metrics associated with a given string must be shown to have diminished such that they are comparable to those of other strings. This change in metrics might be the result of proactive outreach efforts performed by ICANN org or another third party as mentioned above in the section Mitigating Harm. In the case where severe harm threatening human life is suspected were the string to be delegated, then there should be an assurance that that threat no longer exists.

Data Requirements for String Determination
The criteria for determining the state of a potential collision string depend upon the availability of usable and reliable data that can be collected and analyzed by experts. The primary reference data set used in the 2012 analysis by JAS and Interisle was RZM root server data provided by DNS OARC's DITL project. While it is technically possible to repeat the analysis done via JAS and Interisle on current and future DITL data, there is a material concern that the continuously evolving DNS ecosystem changes, via Qname Minimization and other technologies noted in 'Appendix 2 – NCAP Gap Analysis Brief' of the revised Study Two proposal, have significantly impacted or impaired such analysis techniques, such that the name collision assessment results could potentially be unsuitable for risk assessment purposes.[28] There are also other forward-looking concerns, such as the timeliness of DITL's annual collection and the reliance on a third party to collect and make the data readily available for perpetuity, that need to be considered.

[25] JAS report, pg 34
[26] Link TBA, Section 4.1.1 Query Volume Analysis, page 16
[27] https://www.icann.org/en/forms/report-name-collision
[28] https://www.icann.org/en/system/files/files/ssac2021-02-05feb21-en.pdf pg 15

Given the critical dependency of experts relying on the availability of data to make name collision assessments, new data sources or alternative mechanisms to collect data should be required for future evaluations.

Concerns About Data Manipulation

One area of concern involves third-party manipulation of the data used to evaluate the risks associated with name collisions. There are a variety of ways a third party could fabricate the appearance of name collisions in the DNS. At this time, there is no way to predict or prevent this type of manipulation, and identifying the data to differentiate between legitimate name collisions and fabricated ones requires a combination of a quantitative and qualitative data analysis. Moreover, a determined attacker with enough lead time could readily hide the manipulation such that it would be challenging for experts to identify it since such manipulation is both easy and inexpensive. There is also a significant risk here in that with the knowledge that the future name collisions assessors, prospective registrants, or other parties will rely on specific data sources creates an unintended incentive for this manipulation, which could result in very large numbers of unnecessary DNS queries, and thus requiring investigation that might delay name collision analysis by corrupting legitimate data collection mechanisms.

Summary

Undelegated TLD strings can be classified as Collision Strings, and some Collision Strings should be added to a list of strings that should not be delegated. These classifications should be made by using the best information available with the help of technical experts. Qualitative analysis is desirable for the most accurate view, but quantitative analysis using query metrics provides some guidance and due diligence when that is lacking.

Conclusion

The issue of name collisions remains an important concern for the health of the DNS. As noted in the Board’s rationale for its Resolutions of 25 March 2021,

> “The Board's action is expected to have a positive impact on the security, stability and resiliency of the Internet's DNS, as it is designed to continue to study name collisions. This action also serves ICANN's mission in ensuring a secure and stable operation of the Internet's unique identifier systems. This resolution is in the public interest in meeting ICANN's core value of preserving and enhancing the administration of the DNS and the operational stability, reliability, security, global interoperability, resilience, and openness of the DNS and the Internet.”[29]

[29] ICANN Board, “Approved Board Resolutions | Regular Meeting of the ICANN Board | 25 March 2021,” Rationale for Resolutions 2021.03.25.11 – 2021.03.25.14, 25 March 2021,

We have given a definition to name collisions and have described the ways in which they manifest. We have described the harm they might cause and have listed techniques to mitigate such harm. While quantitative approaches are useful for measuring impact and potential harm, they must be accompanied by qualitative analysis to understand the real-world impact of the collision. Policy and implementation choices can reduce risk. Even so, we recognize that no measurement or mitigation technique is comprehensive or completely effective, so these measures reflect due diligence on the part of ICANN org.

It is important to understand that name collisions will not always be observable, even if it is possible for the name collision to exist. There is data that can be collected and can be analyzed, but as will be shown in the complete NCAP Study Two Report, including its appendices, domain names that could manifest a collision can be deployed in private environments and never appear in the collected data.

While the technical aspects of name collision are important to understand, it is best to consider name collision a risk management problem. We are able to define what name collisions are and evaluate some of the root causes, but each scenario must be handled on a case-by-case basis to understand the real-world impact of the collision. The NCAP DG offers guidance on how the ICANN Board might understand and manage the risk in the NCAP Study Two Report.

The NCAP DG expects that the responses to the questions originally posed by the Board will offer guidance as the Board considers the unique risk of each gTLD delegation.

https://www.icann.org/en/board-activities-and-meetings/materials/approved-resolutions-regular-meeting-of-the-icann-board-25-03-2021-en#2.b.