Draft Responses to Board Resolution 2017.11.02.30 by Name Collision Analysis Project Discussion Group

15 January 2024

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). 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 "[Building] 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."

Interpreting Board Questions

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

(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?

(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?

(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

> What factors affect the potential success of the courses of action to mitigate harm?

(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?

(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

> 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?

Through the interpretation exercise, the NCAP DG found themes that tie together various topics. These six (6) themes are used to organize this document:

- 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)

Theme 1: 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.[1] 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.[2] 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 Section 1.2 (Background and Related Work)[3] and Appendix 1 (Revised Definition of Name Collision and Scope of Work)[4] of the Study Two Final Report.

---

[1] See ICANN Name Collision Resources & Information, https://www.icann.org/resources/pages/name-collision-2013-12-06-en
See Name Collision in the DNS ("Interisle Report"), https://www.icann.org/en/system/files/files/name-collision-02aug13-en.pdf
See "Addressing the Consequences of Name Collisions" - ICANN Announcements, 5 August 2013, https://www.icann.org/en/announcements/details/addressing-the-consequences-of-name-collisions-5-8-2013-en
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

[2] 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/en/public-comment/proceeding/proposed-definition-of-name-collisions-and-scope-of-inquiry-for-the-name-collisions-analysis-project-02-07-2019

[3] See Study Two Final Report: Section 1.2 - Background and Related Work

[4] See Study Two Final Report: Appendix 1 - Revised Definition of Name Collision and Scope of Work

Underlying Reasons for Name Collisions

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."[5] The taxonomy consists of six classifications and is thoroughly described in the JAS Report.

The Revised Study Two Proposal for the Name Collision Analysis Project included the following task:

> 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.[6]

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 in the Revised Study Two Proposal.

The Root Cause Analysis examines the documented name collision occurrences reported to ICANN as well as incidents found in Web search results.[7] Per the Root Cause Analysis, the origins of name collisions were diverse, both in terms of the application involved and their mechanisms. 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 namespaces, 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 TLDs is widespread. It is clear from the data that the private use of TLDs is not isolated. Private use of TLDs has been observed for over half of newly delegated TLDs, even though a few TLDs are responsible for more usage than others.

---

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

[6] 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

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

In addition to the Root Cause Analysis, the NCAP DG carried out a research study (Case Study of Collision Strings) that enabled the DG to identify, define, and analyze "Critical Diagnostic Measurements" necessary to sufficiently assess name collision risks. The Case Study of Collision Strings revealed that multiple quantitative measurements taken together were needed to properly assess the scope, impact, and potential harm of name collisions. Nonetheless, currently available data sources and measurement methods limit the ability to analyze root causes of name collisions to identify their likelihood. Access to longitudinal data sources and methods that accurately measure potential for name collision and level of impact are needed to better understand name collisions and their underlying reasons.

Theme 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 role do the negative answers (hereinafter referred to as "negative responses") 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?

As noted in the SSAC Report *Redirection in the Com and Net Domains*, unregistered or unresolvable 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.[8]

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 within the application 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

[8] See *Redirection in the Com and Net Domains*, https://itp.cdn.icann.org/en/files/security-and-stability-advisory-committee-ssac-reports/report-redirection-com-net-09jul04-en.pdf

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.[9] 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.

Another example of how negative answers have a role in system behavior and the end-user experience is the search list.[10] Hosts commonly use search lists for facilitating the resolution of names with common TLDs and other suffixes used by that host. As covered at length in SAC064, the rules are complex and inconsistent from host to host. 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.

This expectation is no longer valid. When the expectation of a negative answer is no longer valid, the end-user experience might be highly variable, ranging from no disruption to complete interruption, without a clear understanding of the cause.

Theme 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 existing users may potentially be exposed to if Collision Strings were to be delegated? What is the likelihood of harm to existing users 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, as well as any other types of harm?

To address the Board's question, the discussion group defined harm as the potential negative impact that might be felt by individuals and organizations experiencing name collisions. There were three aspects of harm the discussion group focused on: potential harm, reported harm, and systemic harm.

[9] ibid

[10] See SAC064: SSAC Advisory on DNS "Search List" Processing (13 Feb 2014), https://itp.cdn.icann.org/en/files/security-and-stability-advisory-committee-ssac-reports/sac-064-en.pdf

Potential harm and reported harm consist of specific instances of name collision. 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 considers the harm to the Internet if users lose trust in the DNS. This type of harm is a broader concern the Board must consider if the risk of name collisions damages the reputation and ability for users to trust the responses for names in the DNS.

In the sections that follow, additional details for each type of harm will be provided, along with specific instances of their impact based on 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.

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 collusion 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 with the resources, knowledge, and skill to identify the root cause of the disruption and implement a remedy or mitigation plan. However, for some users, a response prompting investigation into a potential name collision may be too complicated to understand or remedy without significant technical assistance from a third party. Additionally, this solution may be overly expensive for a large organization. Finally, 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. Even if DNSSEC is deployed, a user may experience either or both communication disruption or 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 historical sources of data from the 2012 round 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.[11] 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."

Systemic Harm

The DNS operates on trust of the integrity of its operation and the validity of its responses. If name collisions become broadly understood as possible anywhere, at any time, with any domain, the risk of harm to the entire DNS is an area the Board must consider in their overall risk assessment of potential name collisions. This concern goes beyond direct harm to users and moves into harm to the system itself.

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. Therefore, careful attention must be provided to understanding causes of varying types and degrees of harm, along with methods for preventing or mitigating harm, at the individual user and system levels.

Theme 4: 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 take that might mitigate harm?

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

(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 presence of potential harm requires possible courses of action that not only mitigate harm but also reduce the likelihood that a negative impact is felt at all. Both reactive and proactive measures can be taken to mitigate the potential for harm associated with name collisions. Both can be effective techniques, but they each also have limitations such that the potential for harm cannot be completely eliminated.

In the sections that follow, we describe possible reactive and proactive harm mitigation courses of action, including which parties might be expected to take action. Additionally, we describe various factors that may affect the outcome of harm mitigation actions.

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."[12] The same report opines that "de-delegation of a TLD in the root would effectively be a permanent death for that TLD." 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 may be deterred from submitting a report unless their " system is suffering demonstrably severe harm … or [they] have a reasonable belief that the name collision presents a clear and present danger to human life".[13] Less severe text would allow for greater capture of name collision data submitted by individual users, 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. This is especially true considering the abundance of knowledge of root causes identified and analyzed in the

[12] See Mitigating the Risk of DNS Namespace Collisions: Final Report (hereinafter referred to as the "JAS Report"), https://www.icann.org/en/system/files/files/name-collision-mitigation-final-28oct15-en.pdf
[13] "Report a Name Collision," ICANN, accessed 17 January 2024, https://www.icann.org/en/forms/report-name-collision

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.[14]

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 Final Report on the new gTLD Subsequent Procedures Policy Development Process ["SubPro Final Report"][15]). There is already precedent for this type of action, as described in the NCAP Study One Report[16]:

> 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

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.[17] 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.

---

[14]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

[15] 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

[16] 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

[17] 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

Proactive Measures to Reduce the Potential for Harm

Equally, if not more, important than sound reaction practices to potential harm are proactive measures that work to prevent or mitigate the impact of harm and to possibly limit the occurrence of events that may cause potential 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. In these cases, the timing is completely unknown, and application-layer data might be exchanged. Thus, controlled interruption potentially causes immediate, short-term difficulties with the intent of preventing greater harm 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.

- As detailed in the Name Collision Analysis Project Study Two Report, methods for notifying users of potential name collision–such as controlled interruption–are crucial for empowering users to remedy or mitigate the possible harm. However, **methods that raise awareness of potential name collisions among impacted parties must be accompanied by sufficient technical assistance and education to be effective.** This is especially important for end-users who may not understand the risks and consequences of name collisions or what steps to take to mitigate potential harm. ICANN will need to continue (and also expand) its education and outreach efforts related to name collisions to support meaningful proactive measures to mitigate harm.

- Another way of proactively reducing the likelihood of harm before it occurs is by **implementing the Name Collision Risk Assessment Framework proposed in the Study Two Report to analyze possible string collisions that carry a potential for harm prior to their delegation**. The NCAP DG in its Recommendations to the ICANN Board affirms the importance of establishing a Technical Review Team function dedicated to assessing characteristics such as high volume, high diversity of queries, and query names under a given TLD, along with other data, on a case-by-case basis.[18] 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. A dedicated Technical Review Team tasked with identifying high-risk strings during the New gTLD Program: Next Round would enable ICANN to limit potential risk of string collisions and mitigate possible harm prior to delegation of the string.
- In some cases, proactive investigation of name collisions by a Technical Review Team might yield a set of TLD strings whose risk of potential harm is significantly high, enough so that it might be prudent to **maintain a Collision String List of potentially high-risk 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, mitigation plans addressed, or the potential harm is cleared. Both identification of high-risk strings and the maintenance of a Collision String List can be done in advance and independently of any TLD-application round.

Factors Affecting Potential Success of Harm Mitigation Measures

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. ICANN, in its mission to help ensure a stable, secure, and unified global Internet, must adopt proactive risk mitigation practices ahead of the *New gTLD Program: Next Round* to prevent possible harm to users and systems related to collision strings. The NCAP DG, in its remit, has provided background information, research studies, recommendations to the ICANN Board, and more in the Name Collision Analysis Project Study Two Report, which are intended to inform ICANN of possible risks and mitigating steps.

[18] "Recommendation 7 - ICANN should establish a dedicated Technical Review Team function," Name Collision Analysis Project Study Two Report

Theme 5: 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?

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 to measure the impact associated with name collisions might not accurately reflect level of risk and potential harm without additional qualitative data and technical review; 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.

Theme 6: 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 whether a Collision String should not be delegated? What are the suggested criteria for determining that 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, if any, of creating a collision string list?

Criteria for Identifying Collision Strings

The NCAP DG in its Study Two Report has proposed a Name Collision Risk Assessment Framework[19] that includes the establishment of a Technical Review Team function to review strings for a high-risk level and place them into the category of "Collision String," which the DG refers to as a *Collision String List*. Among the strings that should be added to the Collision String List due to posing a high level of risk for collisions are .corp, .home, and .mail, which the ICANN Board specifically asked the NCAP DG to research.

In carrying out the Case Study of Collision Strings, the DG has identified various quantitative *Critical Diagnostic Measurements* that can be analyzed along with other data to identify high-risk strings.

To properly identify potential name collisions, a Technical Review Team function should exist to review the data previously mentioned, including name collision data from data sources such as Day-In-The-Life (DITL), Identifier Technology Health Indicators (ITHI) metrics, and ICANN Managed Root Server (IMRS) DNS Magnitude data. This review would need to be done by individuals with significant technical expertise in Internet measurements and the DNS on a case-by-case basis, as described within the Report.

Criteria for Determining Whether a Collision String Should Not Be Delegated

[19] See Name Collision Analysis Project Study Two Report

In addition to providing a proposed process for identifying strings with a high potential for name collisions, the Name Collision Risk Assessment Framework proposed by the NCAP DG includes a process for determining whether a collision string should remain on the Collision String List and not be delegated to an applicant. Since name collisions present risks of various harms to users, systems, and the DNS, assessment of the potential impact of a name collision must be the primary factor in determining whether a string should not be delegated.

As highlighted within the Root Cause Analysis, certain strings may display a high volume of name collisions, but the quantitative measurement of collision volume is insufficient for definitively determining the potential harm and level of impact these strings may have on users, systems, and the DNS. Hence, quantitative measurements identified within the Study Two Report must be balanced with additional data, qualitative measurements, and judgment by a Technical Review Team function.

Following a holistic assessment of potential harm and impact risks posed by a string on the Collision String List, the applicant, the Technical Review Team, and the ICANN org must have an opportunity to reconsider delegation. Strings that display a high risk for high impact and potential harm can remain undelegated to an applicant and kept on the Collision String List until proper mitigation plans or an appropriate remediation effort can be made to neutralize risk. This process is described within the Study Two Report.

Criteria for Determining the Removal of a Collision String from the Collision String List

As described in the previous section, the NCAP DG has proposed a Name Collision Risk Assessment Framework that includes a process for review of high-risk strings on the Collision String List. If, after determining that a string does not pose a high risk of high impact due to collision, based on Critical Diagnostic Measurements and additional data, a string can be moved from the Collision String List back into the application workflow.

In addition to a string being removed from the Collision String List due to diminished metrics associated with its risk of harm, mitigation or remediation plans that reduce potential harm or impact are another criterion for removing a string from the Collision String List. 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 Theme 4: Mitigating Harm.

Measures to Potentially Protect Against Data Manipulation

One area of concern related to the assessment of strings using quantitative Critical Diagnostic Measurements involves third-party manipulation of the data. 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, along with the use of multiple datasets that are published at different

frequencies 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.

To limit the potential manipulation of CDM measurements, reviewers may use longitudinal and historical data as one input to discover aberrant changes. Longitudinal DNS name collision data may need to be captured to improve available measures to detect and protect against data manipulation.

Potential Negative Effects of Creating a Collision String List

The use of a Collision String List for high-risk strings has many advantages for mitigating potential harm caused by name collisions. However, without the necessary resources to appropriately measure risk of collision, harm, or impact, using multiple measurements, including Critical Diagnostic Measurements, DNS name collision data, and qualitative assessment, a string that demonstrates a high volume of collision without additional metrics may be mistakenly added to or kept on a Collision String List despite posing little risk of harm.

Quantitative measurements such as volume query are a critical component of assessing the risk of name collisions, but high volume does not definitively relate to a high level of risk. Additional details and analysis by an expert would be necessary to prevent strings from being inadvertently added to a Collision String List.

This is one potential issue with creating a Collision String List, but with adequate expert reviewers, data, and formalized review processes, ICANN org would be in a position to minimize this potential negative effect.

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.[20]

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 enough 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 that can be analyzed, such as the NCAP Study Two Report and studies, but 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 Recommendations section of 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 delegation in the *New gTLD Program: Next Round* and any future rounds.

[20] 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.