ICANN Org's Office of the Chief Technical Officer (OCTO) appreciates the opportunity to respond to the request for public comments on "Name Collision Analysis Project (NCAP) Study 2 Documents". In this response, we highlight points that were made in each of those documents and suggest updates to the documents. The two documents in this public comment are "Case Study of Collision Strings" ("Case Study") and "A Perspective Study of DNS Queries for Non-Existent Top-Level Domains" ("Perspective Study").

In SSAC2021-02, "Revised Study 2 Proposal for the Name Collision Analysis Project", the two goals of NCAP Study 2 are:

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

The immediately following two sections concern the relationship of the documents to those goals. The final section lists many unsupported conclusions made in the two studies.

Goal #1: Understanding the root cause of most name collisions

"Case Study" does not appear to be related to understanding the root cause of most name collisions. If this case study is meant to be related to that goal, the document should be revised with an explicit description of how the case study relates. OCTO notes that such a relationship would be surprising because queries seen at the root servers cannot be easily traced back to the original stub resolver that caused the query in the recursive resolver. OCTO's understanding is that, without knowing the original context of a query at the stub resolver on the originating system, the root cause of a collision cannot be determined.

Similarly, "Perspective Study" does not appear to be related to understanding the root cause of most name collisions. If this perspective study is meant to be related to that goal, the document should be revised with an explicit description of how the perspective study relates. OCTO again notes that such a relationship would be surprising because queries seen at the root servers cannot be easily traced back to the original stub resolver that caused the query in the recursive resolver, and the original stub resolver context is required to determine root cause.

The root cause of a query could potentially be determined by investigating the circumstances at the stub resolver that caused the query to be sent to the recursive resolver. OCTO is not aware of any public resolver that has attempted such a root cause investigation. Indeed, such an investigation would violate the stated public privacy policies of some of the more popular public resolvers. Even if a public resolver had a policy of logging some queries, it would be stymied because many queries are sent through forwarders. We are aware that many public resolvers are unwilling to share data related to stub resolvers for privacy reasons, so it is not surprising that so little public resolver data was available for the study. If the NCAP DG believes that recursive resolver data is important to research name collision root causes, we suggest that the group work to obtain such data from sources other than public resolvers. For example, could the organizations employing members of the NCAP DG contribute recursive resolver data?

Neither document hints at any method to understand the root cause of most name collisions. OCTO assumes that the NCAP Discussion Group will propose one or more methods in its final report for Study 2, or will say that such methods do not exist. Given the extremely large set of potential name collisions shown in these two documents, OCTO is particularly interested in descriptions of how such methods would find the root causes for many, much less "most", name collisions.

Goal #2: Understanding the impact of name collisions

"Case Study" shows an increasing volume of queries for undelegated TLDs over time, but does not quantify any significant impact of this increased volume on the root server system (RSS). The same is true for the other measurements in the case study: there is increasing diversity among many aspects, but the case study does not quantify any significant impact of any of them on the RSS. The "impact" listed in the goal might be on some other participant in the DNS, such as end users or recursive resolvers, but the query volume and increasing diversity shown in "Case Study" don't appear to relate to those participants either. The document should be revised to specify which, if any, of these increases has a significant impact on the RSS, on end users, or on resolvers, by showing the significance.

Section 5.9 states, *"The study has shown very clear evidence of "impact", e.g., a tremendous amount of query traffic would be affected by a delegation of .CORP, .HOME, and .MAIL."* This statement is not supported by any evidence in the document, and in fact may be contradicted by recent delegations of new gTLDs. For example, an informal review of IMRS traffic saw very little impact to the RSS after .MUSIC was delegated in November 2021, and there has been no public indication of any impact on resolvers by the delegation. Given this finding by OCTO, this document should be updated with the "very clear evidence" or the statement should be removed.

Similarly, "Perspective Study" does not quantify any significant impact of undelegated TLDs on the RSS, on end users, or on resolvers. In the section comparing the names seen at a public recursive resolver and the RSS, there is also no quantitative analysis of the impact of the rate or diversity of undelegated TLDs on this unnamed public resolver or any other resolvers. The document should be revised to specify which, if any, of the data shows any significant impact on the public resolver used or other resolvers.

Unsupported conclusions in the Case Study and Perspective Study

Both “Case Study” and “Perspective Study” contain numerous conclusions that are not supported by the data reported in the documents.

The unsupported conclusions in “Case Study” include:

- Executive Summary: *“These facts suggest that challenges relating to impact and mitigation are also increasing.”*
  None of the increases in the case study are shown to be challenges.
- Section 5.1: *“This is because the expectation of negative responses is high, and the mitigation across multiple services, networks, and users is increasingly complex to perform.*”
  The case study has no qualitative or quantitative review of controlled interruption, the form of mitigation used on over 1000 gTLDs since 2013. Without such review, the phrase “increasingly complex” cannot be supported.
- Section 5.2: *“The sheer volume of query traffic for the undelegated names under study is alarming in itself, ...”.*
  The case study does not support any reason for alarm. If the RSOs indicate any concern, that should be reflected in the document; otherwise, the document needs to show why the impact is significant to resolvers or users.
- Section 5.3: *“This finding highlights the challenge associated with mitigation since diversity complicates mitigation coordination across an increasing number of parties (i.e. networks, vendors, applications, and users).”*
  Without a qualitative or quantitative review of the current use of controlled interruption, the word “challenge” cannot be supported.
- Section 5.4: *“This finding also highlights the challenge associated with mitigation since diversity complicates mitigation coordination across more systems, applications, etc.”.*
  Without a qualitative or quantitative review of the current use of controlled interruption, the word “challenge” cannot be supported.
- Section 5.5: *“This information can also help construct mitigation strategies.”*
  This statement is unsupported because no mitigation strategies are suggested here.
- Section 5.8: *“This diversity poses the greatest challenge for mitigation since there are potentially countless ramifications to be identified and resolved.”*
  Without a qualitative or quantitative review of the current use of controlled interruption, the word “challenge” cannot be supported.
- Section 6: *“The analysis illuminates the significant impact delegations would mean, and presents some insight into the potential harm that may result.”*
  The analysis does not show how much significance the impact of delegations has on the RSS, on end users, or on resolvers.

The unsupported conclusions in “Perspective Study” include:

- Executive Summary: *“These findings are significant in terms of how future guidance and advice should be applied to name collision risk assessments.”*
  The study has no risk analysis or data about delegating, or continuing to not delegate, currently undelegated TLDs.
- Section titled “Non-Existent TLDs with Highest Query Count”: *“This helps support and inform the DNS community that the publication of top-N strings could be beneficial to future TLD applicants.”*
  The study does not discuss future TLD applicants, much less any possible benefits to them of knowing which strings might have more or fewer collisions seen in the RSS.
- Section titled “Key Findings”: *“The two studies in this analysis provide two key findings that will help the NCAP provide guidance and advice to ICANN as to how future risk assessments of name collision strings should be evaluated.”*
  The study does not discuss any risks of current or future name collisions. The implications listed in the rest of this section have the same problem with assuming that any risk has been shown or needs to be studied.

The documents should be revised to ensure that all the stated conclusions are supported by the data included in the documents.

OCTO looks forward to the revised documents, and also looks forward to the final report for Study 2.