NCAP Gap Analysis Brief

This document serves as a short brief that describes both the perceived technical and data gaps that were identified by the NCAP discussion group and should be considered as inputs to help form investigative research tasks for subsequent NCAP Studies. Those studies will incorporate these considerations and potential data sources that were not utilized to quantitatively or qualitatively assess name collision risks in the 2012 program and help provide guidance to ICANN Board's questions in their resolution 2017.11.02-29 – 2017.11.02.30. The items below are loosely codified into four areas and include supporting annotations to help illustrate where they apply to each subsequent study and how they apply to their specifically relevant Board questions and help answer them.

Background

The Study 1 report of the Name Collision Analysis Project (NCAP) provides a concrete definition of the term "name collision" and serves as a summary report on the topic in which it brings forth important knowledge from prior work in the area (directly addresses Board Question 1). While some name collision research was conducted in the years prior to the new gTLD program in 2012, the field was and still remains an esoteric field of cybersecurity research. However, since the last risk assessment of the new gTLD program was conducted, peer reviewed academic proceedings and industry reports have been published that highlight the more nuanced concerns, threats, vulnerabilities, and underlying causes of name collisions within the DNS. Furthermore, the internet's DNS ecosystem has evolved since the previous round of TLD delegations to a state in which there is more name server consolidation as well as protocol bifurcation and alterations that may directly impair the observational capacity to conduct name collision risk assessments. To that end, we believe there is a gap of substantive data resources and knowledge between the 2012 round and now that should be considered when assessing the risk profile and mitigating controls to deploy for future TLD delegations by ICANN.

1.) Changes in the DNS infrastructure and Protocol: DNS usage monitoring provides insight into time-resolved traffic evolution patterns useful in the quantification of system stability and performance as well as detecting aberrant events. Longitudinal measurements and usage trends, however, are increasingly difficult to leverage as the underlying system evolves or as bifurcation within the system occurs. These system changes may result in non-symmetric system usage, partial or even total impairments in DNS measurements, and ultimately confound the interpretability of the system's usage metrics. Since the 2012 round of TLD delegations, several new technologies and recommended best practices within the DNS ecosystem now have a significant impact on the volume and fidelity of DNS queries observed at name servers in the DNS hierarchy. These technologies include running Root on Loopback (RFC 7706), Aggressive Use of DNSSEC-Validated Cache (RFC 8198), DNS Query Name Minimization (RFC 7816), and DNS Queries over HTTPS (RFC 8484). It is in the DNS community's best interest to develop a better understanding of how these standards and technology changes will influence data collection capabilities as well as their impacts to data analysis of DNS traffic in an ever evolving, technologically fragmented, and highly distributed system (Board Questions 2 and 7 - Study 2).

2.) Controlled Interruption Efficacy and Data Analysis: While the NCAP Study 1 Report highlights some reports around the efficacy of Controlled Interruption, we believe a more thorough assessment of the framework should be commenced. The collected reports (which comprises a data set previously unavailable for study) should at a minimum be analyzed to better understand any trends, commonalities, assumptions, and success attributes (Board Question 4 - NCAP Study 2). Understanding the nature of these reports with a re-examination of previous Day In The Life (DITL) of the internet data may help identify key signals in the DNS that could better inform name collision risk assessments moving forward (Board Questions 5, 6, and 8 - NCAP Study 2). Some applications, including popular browsers, have implemented specific DNS controls to signal when Controlled Interruption events occur. To that end, efforts should be made to identify and contact such vendors to see if instrumentation data is available. Finally, a study should be made to provide additional evidence that Controlled Interruption was a successful mitigation model, which may include creating and running simulation test beds (Board Questions 4, 5, and - NCAP Study 3).

3.) Vulnerability Understanding and Mitigation Strategies: Since the 2012 delegation of TLDs, various peer reviewed academic and industry papers have been published that elucidate some of the more detailed nuances of name collisions, specifically as they relate to various potential risks and vulnerabilities (Board Questions 3, 7, and 8 - Study 2). Specifically, many of these publications directly identify known DNS query patterns, typically associated with zero-configuration protocols such as DNS-SD, that we believe may be weaponized and exploited in a name collision environment. If true, this new knowledge should be applied to future TLD delegation risk assessments as it builds upon a foundational understanding of the intent of the DNS queries as opposed to the volume of queries that was originally used in the new gTLD risk assessment (Board Questions 4, 5, and 6 - Study 3).

4.) Data Sets: Since the new gTLD program, various new data sets have become available that may provide additional telemetry to better understand and assess name collision risks. The new gTLD name collision risk assessment was conducted against a few years of DITL DNS traffic data. Unfortunately, the DITL data set has several limitations, as it only provides a few days per year of authoritative root server DNS traffic, is contributed by root server operators on a voluntary basis, may be anonymized due to privacy concerns, and as noted in Item 1 above may require a different method of analysis. Since the 2012 TLD round of delegations, the collection of DITL data has continued and may provide better longitudinal measurements pre/post the new TLD delegations. Other entities have also started to retain high fidelity root DNS traffic that may provide better insights. The emergence of popular open recursive resolvers has also transpired and dramatically shaped the DNS ecosystem since the new gTLD delegations. These recursive services may provide a richer and more complete understanding of name collisions if they can be utilized for analysis. Other potential data repositories of interest would also include the ORDINAL DNS data as well as Certificate Transparency records, neither of which existed during the previous assessment.

Appendix 1 - Relevant Board Questions

(1) a proper definition for name collision and 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;

(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;

(4) possible courses of action that might mitigate harm;

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

(6) potential residual risks of delegating Collision Strings even after taking actions 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;

(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

(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.