INTERNET-DRAFT

Intended Status: Informational

Expires: August 6, 2015

Matthew Thomas Allison Mankin Verisign Labs Lixia Zhang UCLA February 6, 2015

Report from the Workshop and Prize on Root Causes and Mitigation of Name Collisions draft-thomas-namecollisions-workshop-report-00

Abstract

This document provides context and a report of a workshop on Root Causes and Mitigation of Name Collisions, which took place in London, United Kindgdom, on March 8 to 10, 2014. The main goal of the workshop was to foster a discussion on the causes and potential mitigations of domain name collisions. This report provides a small amount of background and context, then provides a summary of the workshop's discussions.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of <u>BCP 78</u> and <u>BCP 79</u>.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

Copyright and License Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to <u>BCP 78</u> and the IETF Trust's Legal Provisions Relating to IETF Documents

(http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Table of Contents

<u>1</u> Introduction	. <u>3</u>
2 Background and Context	. 4
<pre>2.1 Brief Update</pre>	. 7
3 Workshop Structure	. 7
3.1 Research Findings	. 8
3.2 System Analysis	. 9
3.3 Frameworks - Modeling, Analysis & Mitigation	. 9
3.4 Conclusions and Next Steps	. 10
4 Security Considerations	. 11
5 IANA Considerations	. 11
6 Acknowledgments	
<pre>1 Informative References</pre>	. 12
Appendix A. Program Committee	
Appendix B. Workshop Material	. 14
<u>Appendix C</u> . Peer-Reviewed Name Collision Papers	. 14
<u>Appendix D</u> . Invited Name Collision Talks	. 14
<u>Appendix E</u> . Panels and Discussions	. 15
<u>Appendix F</u> . Workshop Participants	. 15
Authors' Addresses	. 15

1 Introduction

It has been well known within the Internet research and engineering community that many installed systems in the Internet query the DNS root for a wide range of top-level domains (TLDs). Many of these TLDs are not delegated which results in a response indicating that the name queried does not exist (commonly called an NXDOMAIN response). In the ICANN community, it was observed as early as November 2010 by the Security and Stability Advisory Committee (SSAC) report [SAC045] that the addition of new top-level domains in the DNS root could result in so-called name collisions for installed systems. Some installed systems, following established (albeit not vetted) operational practices, generate queries to the global DNS with name suffixes that under seemingly reasonable assumptions at the time the systems were designed or configured, were not expected to be delegated as TLDs. Many of these installed systems depend explicitly or implicitly on the indication from the global DNS that the domain name suffix does not exist. After a new TLD is delegated, the global DNS may give a different response to the query involving the TLD than it did prior to the TLD's delegation.

A name collision occurs when an attempt to resolve a name used in a private namespace results in a query to the public Domain Name System (DNS), and the response indicates that the name is in the global DNS [NCRI]. In other words, the overlap of public and private namespaces may result in potential unintended (and therefore potentially harmful) resolution results. The impact of the global change on installed systems will be varied; risks to installed systems introduced by name collisions may arise due to varied causes.

In a globally distributed system such as the Internet, it is difficult yet critical to enforce policies for demarking boundaries of ownership and autonomy. The Internet's global DNS is an exemplary system in which name space governance is critical to ensure name uniqueness.

In order to help ensure this uniqueness and interoperability, several namespaces are administered by ICANN for the global global benefit of Internet users. Prior to ICANN's creation in 1998, seven generic Top-Level-Domains (TLDs) were defined in the early development of the Internet [RFC1591]. Since the formationof ICANN, the delegations of generic, internationalized and country code TLDs have been administered and delegated by ICANN. During these delegations, it quickly became apparent within the IETFcommunity there was a need to reserve name spaces that can be used for creating limited sets of internal names without fear of conflicts with current or future TLD name spaces in the global DNS [RFC2606].

While [RFC2606] aimed to enable operators to use these only a small set of reserved names internally, with limited uses, educational awareness and operational best practices did not achieve the goal [RFC 6761], and other suffixes, not reserved though at the time not in conflict, were often employed instead. Faulty assumptions of "we only use this name internally and there is no possibility of leakage to the global DNS" were made by numerous operators or administrators. Numerous reports and findings have clearly disproved these faulty assumptions by showing substantial "DNS leakage" into the global DNS through mechanisms such as search lists.

In 2012, ICANN created a new gTLD program to add a potentially unlimited number of new qTLDs to the root zone as a mechanism to enhance competition, innovation and consumer choice. With the potential of many new qTLDs becoming delegated in the global DNS, operators or administrators who elected to use the same name space internally may face potential "name collision" problems.

This document is primarily a report on the March 2014 workshop that set out to examine the causes and mitigation of such name collisions and their associated risks. It is a companion to the workshop proceedings at [WPNC], and it also provides some additional background and context.

2 Background and Context

When the workshop was convened the context and status of the work around name collisions could be described as follows.

Since early 2008, there had been numerous lengthy discussions within the ICANN community about the ability of the DNS root to scale to accommodate new gTLDs, and the impact of making those changes on the DNS ecosystem. In March 2008, the Internet Architecture Board observed that the introduction of suffixes in use in a number of environments could lead to instability [IAB2008]. The Security and Stability Advisory Committee (SSAC) in December 2010 issued [SAC046] in which the committee formalized several recommendations based on "actual measurement, monitoring, and data-sharing capabilities of root zone performance" to help determine the feasibility of root scaling. Separately, the Root Server System Advisory Committee [RSSAC] agreed in late 2010 on the need to establish standard metrics to be collected and reported by all operators. This effort would provide the community with a baseline measure of the entire root server system's performance. With such an established baseline, any possible negative effect from additional TLDs within the root could potentially be identified. In late 2012, the ICANN Board affirmed the Thomas

need to work with the root server operators via RSSAC to complete the documentation of the interactions between ICANN and the root server operators with respect to root zone scaling [IR2012].

In March 2013, SSAC published an advisory titled "SSAC Advisory on Internal Name Certificates" [SAC057], which identified a CertificateAuthority (CA) practice that, if widely exploited, "could pose a significant risk to the privacy and integrity of secure Internet communications". The ICANN Board acknowledged the issues identified in [SAC057] as part of a more general category of issues. These issues included installed systems utilizing a namespace in a private network that includes a non-delegated TLD that is later delegated into the root. In May 2013, the ICANN Board commissioned a study on the use within private name spaces of TLDs that are not currently delegated at the root level of the global DNS [ISTUDY]. This study was focused on potential name collision events between applied-for new gTLDs and non-delegated TLDs potentially used in private namespaces. The study also examined the potential possibility of name collisions arising from the use of digital certificates referenced in [SAC057].

Between the [RSSAC] and [SAC046] advisory statements and the ICANN commissioning of a study in May 2013, there was significant progress on establishing coordinated monitoring and measurement of the root. RSSAC approached its finalization of the specific metrics that each root operator will collect and initiated discussions about where the operators will send their data for analysis once collected. To properly gauge the risks of new gTLD delegations to the root, an established baseline of normal performance of the system would be required to start sufficiently ahead of the new delegations. The timing and execution of these RSSAC and SSAC recommendations timed poorly with the commissioned study, resulting in a limited pool of data repositories from which any baseline and risk measurements could be established.

It is common practice for each operator to monitor its own root server, and some operators report the status and performance of their services publicly. As of ICANN's study commissioned in May 2013 [ISTUDY], there was no mechanism in place to allow a detailed view of the entire root system, short of the annual "Day in the Life" ([DITL]) data repository, which contains root DNS data over a short coordinated time period from a varying subset of root operators and was intended to be used for research purposes, not to provide overall monitoring and an operational view of system health. Due to the lack of a more comprehensive and desirable data repository for baseline and collision analysis [DITL] became the de facto referential dataset for root traffic analysis. There still remains no holistic view of the root system that would enable the technical community to

determine the potential impact of added TLDs.

The commissioned study, conducted by the Interisle Consulting Group, was published in August of 2013. The report "Name Collisions in the DNS" [INTERISLE], based on [DITL] measurements, addressed name collisions in the DNS and also recommended options to mitigate the various name collision risks. The study identified categories of strings according to the risk they represent: low risk (80% of applied-for strings), uncalculated risk (20% of applied-for strings), and high risk (2 applied-for strings).

At the same time as the [INTERISLE] study, ICANN published a proposal, titled "New gTLD Collision Occurrence Management Plan" [NGCOMP], to manage the risk of name collisions within the applied-for gTLDs. Based on [INTERISLE] measurements, ICANN deemed two strings, .home and .corp, as high-risk because of their widespread use within internal networks and would indefinitely delay their delegation. Those strings within the uncalculated-risk classification would be delayed 2-3 months in their application process while ICANN conducted more research into whether the string is of high or low risk classification. Those in the low-risk classification would face a delay in activating domains until 120 days after contracting with ICANN to allow for the change in certificate authority practices recommended in [SAC057].

Within the ICANN proposal, an approach termed the "alternative path to delegation" was outlined, in which a registry operator could elect the option to proceed with delegation provided it initially blocked all second-level domains (SLDs) that appeared in the certain DITL datasets pending the completion of the assessment. The majority of new gTLD applicants that were eligible elected this alternative path once otherwise approved for delegation. The plan also outlined an outreach campaign to educate system administrators, software developers, and other engineers about the name collision issue and possible mitigation measures.

As a further provision, the "New gTLD Collision Occurrence Management Plan" called for a follow-up study that would develop a "Name Collision Occurrence Management Framework" [NCOMF]. In February 2014, the document "Mitigating the Risk of DNS Namespace Collisions: Phase One Report" was published by the ICANN-contracted group JAS Global Advisors [MRDNC]. The report provides a number of recommendations for addressing the name collision issue focusing on a technique termed "controlled interruption", in which a registry would temporarily resolve all second-level domains (or all second-level domains present in the block list) to a specific IP, 127.0.53.53. The report also makes provisions to implement an emergency plan and strategy in case name collisions had a "clear danger to human life".

2.1 Brief Update

In the timeframe after the workshop, a final version of the Phase One Report was released in June 2014 [MRDNC].

In July 2014, after a community review phase, a final recommendation was issued by ICANN [NCOMFINAL] and this has been followed by the publication of managment documents for the implementation of a controlled interrupt for new gTLD delegations [NOCA], [NCSLDCIV], [ADDNOCA].

One part of the original framework was not released by the time of this document and [MRDNC] indicated that its publication was delayed due to a security vulnerability identified during the course of the work.

Broad community efforts to measure the impact of name collisions were not included in [NCOMFINAL]. At the time of this writing, RSSAC has just published its specification of common measurements to be collected by root operators, meeting one part of the needs [RSSAC002].

3 Workshop Structure

The Workshop and Prize on 'Root Causes and Mitigation of Name Collisions'[WPNC], sponsored by Verisign, took place March 8-10, 2014 in London, United Kingdom. The WPNC was open to the public, and it gathered subject area specialists, researchers and practitioners to discuss and present their views, concerns and ideas surrounding the name collision issue. Proceedings are published at [WPNC].

The workshop focused on studies of name collision risks and mitigations with the expectation to advance the global community's insight into operational uses of name suffixes that can result in name collisions, and to gain a stronger understanding of the potential risks for the users of the installed systems. Additional emphasis and attention was given to discussions that may advance the state of knowledge about the architecture and impacts of DNS namespaces with multiple scopes or resolution contexts and the utilization of new methods of monitoring and understanding the needs and methods for mitigating emerging Internet risks around name collisions. A technical program committee, whose members spanned a variety of organizations and universities, was assembled. The committee issued a call for papers and evaluated all submissions to ensure the highest level of quality.

A synthesis of the accepted papers and conference proceedings is

[Page 7]

Thomas

captured in the subsections below. An informal synopsis of the workshop combined with individual statements and observations is available online [COMMENTARY].

3.1 Research Findings

Many of the research papers focused on the analysis of DITL data to better understand various aspects of the root NXDomain traffic[TECHNIQUES], [RARDBITS], [BLOCKLISTS], [MODELING], [SEARCHLISTS]. Note: all workshop contributions are listed in Appendices C, D, and E, and full papers and slides are available at [WPNC].

While the DITL data has become the de facto referential dataset for root traffic analysis, some presenters echoed concerns that the dataset may have become biased or polluted with "artificial" queries after the ICANN "Reveal Day", in which the list of applied-for gTLD strings was publicy disclosed. No conclusive or empirical evidence of tampering was presented; however, concerns about the integrity and reliability of future DITL collections and analysis for purposes related to new gTLDs were echoed by some panelists [IESCPANEL]. Furthermore, the statistical accuracy and completeness of DITL data used to draw inferential conclusions or more specifically create second-level domain (SLD) block lists - was examined. The efficacy of blocking domains based on sampled DNS data, e.g. DITL, was investigated by comparing measurements of SLDs within DITL and that of a multi-month root NXDomain collection at the A and J roots [BLOCKLISTS]. The findings provided useful insights into SLD-root affinities, SLD temporal query patterns and occurrence frequencies that demonstrated the ineffectiveness of block listing domains based on sampled DNS data such as [DITL].

Measurements of queries specifying the recursion desired (RD) bit to the roots in DITL were quantified to identify the level and nature of naive DNS clients and to determine and assess potential impacts that could arise from the proposed SLD blocking technique to these naive clients [RARDBITS]. A substantial proportion of the root server request traffic contained queries with the RD bit specified. Both in absolute and relative terms, requests specifying the RD bit for applied-for qTLD were found to be significantly lower when compared to existing TLDs. The root cause determination of what system or mechanism is responsible for generating the queries was inconclusive and only speculative explanations of faulty implementations of a DNS resolving server were hypothesized. However, the analysis was also not able to identify instances of actual or potential harm resulting from these naive clients, suggesting if SLD blocking techniques were to be utilized it is unlikely there would be any negative impact to

these naive clients.

3.2 System Analysis

Comparison of elements can often help us to understand a system as a whole. A passive study of the DNS traffic in a provisioned domain such as "corp.com." may elucidate certain name collision parallels [CORPCOM]. Such measurements were presented as a proxy for the ".corp" potential new gTLD. According to the study, significant DNS traffic volume was directed at a variety of third-level domains under "corp.com". This prompted a series of questions surrounding how name collisions can be identified, as most end-users won't recognize that problems may be due a name collision. How will users know that the problem they're experiencing is a result of a new, colliding gTLD? Will support groups be able to diagnose a name collision event from reported symptom(s)? Will a collision-based security hole be detectable?

These questions, whose underpinnings rely on communication and educational awareness, may find recommendations or parallels from other system references during the workshop [JASFRAMEWORK] - such as the postal and telephone system. Most telephone and postal systems have evolved over time, requiring individuals to alter the way they address their parcels or place their calls. Both systems implemented their changes in such a way that prior to the change, educational material is distributed and communicated and for a period of time and after the change compliance of the previous standard is temporarily accepted. While the telephone and postal system operate in a very different way than the DNS, these parallels of "advanced notification, education and communication, and a grace period" were insightful for how other similar systems transitioned.

3.3 Frameworks - Modeling, Analysis & Mitigation

Statements from several TLD operators during the conference reverberated a theme for the need of improved tooling, education and communication surrounding name collisions. The delegation of new gTLDs is an ongoing event and there is a clear and immediate need for these operators to have visibility to monitor and measure the effects of these new gTLD delegations. A lack of tools, data, communication and education surrounding name collisions has handicapped operators in their ability to quantitatively measure and proactively provide any steps for mitigation of risks. To this end, numerous techniques, frameworks and models that focused on the concepts of analyzing,

detecting, and measuring various name collision risk factors were presented and reviewed in hopes to understand these underlying concerns and issues [TECHNIQUES], [MODELING], [SEARCHLISTS], [DNSENDUSER], [ENTNETWORK].

Data-driven analysis and mitigation require operators to be versed and skilled with data analysis techniques to better understand the contextual intent and ownership of DNS queries. An overview of various DNS analysis techniques in which ways of decomposing names, measuring temporal distributions between queries and detecting organizational/geographical affinities, was presented [TECHNIQUES]. More specific techniques were also showcased - such as a systematic way of observing and characterizing the impact of search lists within root DNS traffic allowing operators to quantify the number of unique entities that may be reliant on a particular name space [SEARCHLISTS]. While not exhaustive, the techniques presented have been proven to elucidate patterns within root DNS traffic data and could serve as the potential building blocks of a DNS analysis framework.

Most of the previously published work focused on name collisions has produced various quantitative analyses based on observations of Internet traffic and data, including DNS queries and Web content, in which behavior and associated risks have been inferred. An understanding of the inverse of the process by starting with a fundamental model of name resolution at the client was proposed as an alternative means to define risk [MODELING]. This model deconstructed the process of name resolution at the resolver library of a client system and formalized a model from which derived metrics could be used to define and quantify associated risks. While the model presented is only a piece of the greater name collision puzzle, it provides potentially new insights into otherwise what may be considered a missing piece.

Just as important as understanding the root causes of name collisions, providing effective mitigation strategies is a critical piece of the name collision puzzle. Mitigation can be achieved from both higher levels, such as ICANN, as well as the enterprise level. Proposed strategies for mitigating name collisions at both of these levels were presented. While the technical details for each proposed strategy varies, underlying dependencies in both strategies require operators to monitor and educate/train their users.

3.4 Conclusions and Next Steps

In their concluding statement, the workshop committee stated: "It occurs to the program committee that the analysis of the interactions

between the different uses of domain names within local of global context is almost a non-existent topic of research. This may have to do with the lack or accessible data, lack of theory of root causes, or a lack of interest, or a bias in the participation of the workshop. We think that this is evidence that this study of the global centrally important technical system needs to be ramped up".

Follow-on commentary [NEXTSTEPS] from the attendees reaffirmed this opinion with recurring messages of a need to understand the root causes of name collision and the need to overcome shortcomings within our data collection, monitoring and analysis of the DNS.

Many name collision unknowns still exist. What are the root-causes of these queries? What is going on within a recursive name server? What vulnerabilities or subtle attack vectors do these new gTLD delegations enable? The limited datasets available to researchers operators are not sufficient to draw baseline measurements for these questions - forcing the community to make inferences and rank guesses as to what is going on within the DNS. Using these sub-optimal data repositories to create solutions such as block lists are only dealing with the symptoms of the problem and not addressing the root cause. To properly answer these questions, the community needs to address the issue of a shortage of funding and data collection / analysis. Communication and educational outreach programs need to be improved in order raise the awareness of impacted parties and broadening the participation and sharing.

4 Security Considerations

Workshop participants discussed security aspects related to root cause analysis and mitigation techniques of potential name collision events. As noted in several papers and presentations, security concerns both may arise, and may be addressed, with name collision mitigation techniques. Follow-on measurement-based research is important to security considerations for name collisions.

5 IANA Considerations

This memo includes no request to IANA.

6 Acknowledgments

We would like to thank both the program committee (Appendix A) and the workshop participants (Appendix F), with equal appreciation to those who spoke formally and those who joined in the lively

discussions.

Additionally, we would like to thank the following persons for their review comments: Burt Kaliski, Olaf Kolkman, and Danny McPherson.

7 Informative References

- [RFC2606] D. Eastlake, A. Panitz, "Reserved Top Level DNS Names", RFC 2606 (also BCP 32), June 1999.
- [RFC6761] S. Cheshire, M. Krochmal, "Special-Use Domain Names", <u>RFC</u> 6761, February 2013.
- [WPNC] "Workshop and Prize on Root Causes and Mitigation of Name Collisions (WPNC)", http://namecollisions.net/, June 2014.

- [SAC057] "Advisory on Internal Name Certificates", http://www.icann.org/en/groups/ssac/documents/sac-057-en.pdf, March, 2013.
- [ISTUDY] "Security Studies on the Use of Non-Delegated TLDs, and

Dotless Names", https://www.icann.org/en/news/announcements/announcement-28may13-en.htm, May 2013.

- [DITL] "A Day in the Life of the Internet (DITL)" , http://www.caida.org/projects/ditl/, July 2011.
- [INTERISLE] "Name Collision in the DNS", https://www.icann.org/en/about/staff/security/ssr/namecollision-02auq13-en.pdf, August 2013.
- [IAB2008] "The IAB's response to ICANN's solicitation on DNS stability", https://www.iab.org/documents/correspondencereports-documents/docs2008/2008-03-07-icann-new-gtlds/, March 2008.
- [NGCOMP] "New gTLD Collision Risk Mitigation", https://www.icann.org/en/about/staff/security/ssr/newgtld-collision-mitigation-05aug13-en.pdf, August 2013.
- [NCOMF] "ICANN Selects Lead for Development of Name Collision Occurrence Management Framework", http://www.icann.org/en/news/announcements/announcement-2-11nov13-en.htm, November 2013.
- [NCOMFINAL] https://www.icann.org/en/system/files/files/namecollision-framework-30jul14-en.pdf
- [MRDNC] "Mitigating the Risk of DNS Namespace Collisions", https://www.icann.org/en/about/staff/security/ssr/namecollision-mitigation-26feb14-en.pdf, February 2014.
- [RSSAC002] "Advisory on Measurements of the Root Server System", https://www.icann.org/en/system/files/files/rssac-002measurements-root-20nov14-en.pdf
- [COMMENTARY] "Proceedings of Name Collisions Workshop Available" , http://www.circleid.com/posts/20140326 proceedings of name collisions workshop available/, March 2014.
- [NOCA] "Name Collision Occurrence Assessment", http://newgtlds.icann.org /sites/default/files/agreements/name-collision-assessment-04aug14-en.htm , August 2014.
- [NCSLDCIV] "Name Collision SLD Controlled Interruption Variations" , http://newgtlds.icann.org/ sites/default/files/agreements/name-collision-sld-

controlled-interruption-12sep14-en.htm , September 2014.

[ADDNOCA] "Addendum To Name Collision Occurrence Assessment", http://newqtlds.icann.org/ sites/default/files/agreements/name-collision-assessmentaddendum-14nov14-en.htm , November 2014.

Appendix A. Program Committee

This workshop program committee consisted of Geoff Huston, Burt Kaliski, Olaf Kolkman, John Levine, Allison Mankin, Lixia Zhang, Anne-Marie Eklund Loewinder, and Andrew Sullivan.

Appendix B. Workshop Material

Main Workshop Page: http://namecollisions.net/

Name Collision Invited and Submitted Papers, Panels and Videos: http://namecollisions.net/program/index.html

Appendix C. Peer-Reviewed Name Collision Papers

- [TECHNIQUES] "Analysis Techniques for Determining Cause and Ownership of DNS Queries" by Matthew Thomas and Andrew Simpson
- [RARDBITS] "Analysing the Use of the RA and RD bits in Queries to Root Servers" by Jim Reid
- [BLOCKLISTS] "The Effectiveness of Block Lists in Preventing Collisions" by Matthew Thomas, Yannis Labrou and Andrew Simpson
- [MODELING] "What's in a Name (Collision): Modeling and Quantifying Collision Potential" by Casey Deccio and Duane Wessels
- [SEARCHLISTS] "Detecting Search Lists in Authoritative DNS" by Andrew Simpson

Appendix D. Invited Name Collision Talks

[KEEPEYE] "Keeping an Eye on Name Collisions" by Bruce Schneier

[CORPCOM] "Looking at corp.com as a proxy for .corp" by Colin Strutt

[DNSENDUSER] "Measuring DNS Behaviors from the End User Perspective" by Geoff Huston

[DNS-OARC] "DNS-OARC" by Keith Mitchell

[ENTNETWORK] "Name Collision Mitigation for Enterprise Networks" by Paul Hoffman

Appendix E. Panels and Discussions

[IESCPANEL] "Internet Engineering and Standards Considerations" by Suzanne Woolf, Peter Koch, Olaf Kolkman, Warren Kumari, and John Levine

[JASFRAMEWORK] "Name Collisions Management Framework" by Jeff Schmidt

[NEXTSTEPS] "Workshop Wrap-Up and Next Steps" by Burt Kaliski

<u>Appendix F</u>. Workshop Participants

A list of workshop participants is provided at [WPNC].

Authors' Addresses

Matthew Thomas

Email: mthomas@verisign.com

Allison Mankin

Email: amankin@verisign.com

Lixia Zhang

Email: lixia@cs.ucla.edu